

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2001, 18:30:40 ; Search time 1981.17 Seconds  
(without alignments)  
95.427 Million cell updates/sec

Title: US-09-675-650-4

Perfect score: 20  
Sequence: 1 gagtagaagagatagaacg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1022815 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_esthum29:\*  
64: em\_esthum30:\*  
65: em\_esthum31:\*  
66: em\_esthum32:\*  
67: em\_esthum33:\*  
68: em\_esthum34:\*  
69: em\_esthum35:\*  
70: em\_esthum36:\*  
71: em\_esthum37:\*  
72: em\_esthum38:\*  
73: em\_esthum39:\*  
74: em\_esthum40:\*  
75: em\_esthum41:\*  
76: em\_esthum42:\*  
77: em\_esthum43:\*  
78: em\_esthum44:\*  
79: em\_esthum45:\*  
80: em\_esthum46:\*  
81: em\_esthum47:\*  
82: em\_esthum48:\*  
83: em\_esthum49:\*  
84: em\_esthum50:\*  
85: em\_esthum51:\*  
86: em\_esthum52:\*  
87: em\_esthum53:\*  
88: em\_esthum54:\*  
89: em\_esthum55:\*  
90: em\_esthum56:\*  
91: em\_esthum57:\*  
92: em\_esthum58:\*  
93: em\_esthum59:\*  
94: em\_esthum60:\*  
95: em\_esthum61:\*  
96: em\_esthum62:\*  
97: em\_esthum63:\*  
98: em\_esthum64:\*  
99: em\_esthum65:\*  
100: em\_esthum66:\*  
101: em\_esthum67:\*  
102: em\_esthum68:\*  
103: em\_esthum69:\*  
104: em\_esthum70:\*  
105: em\_esthum71:\*  
106: em\_esthum72:\*  
107: em\_esthum73:\*  
108: em\_esthum74:\*  
109: em\_esthum75:\*  
110: em\_esthum76:\*  
111: em\_esthum77:\*  
112: em\_esthum78:\*  
113: em\_esthum79:\*  
114: em\_esthum80:\*  
115: em\_esthum81:\*  
116: em\_esthum82:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	16.8	84.0	405	123	AW982117 SSS0027 S
C 2	16.8	84.0	411	141	BE859183 SSS0038 S
C 3	16.8	84.0	511	140	BE810166 SSS038 S
C 4	16.8	84.0	246	146	AZ609094 1M0433L11
C 5	16.8	84.0	619	104	AJ395794 AJ395794
C 6	16.4	82.0	196	167	BE448549 BE448549
C 7	16.4	82.0	207	114	AW311691 AW311691
C 8	16.4	82.0	231	138	BE648384 UT-M-BH2
C 9	16.4	82.0	261	30	AV368821 AV368821
C 10	16.4	82.0	266	125	BB104945 BB104945
C 11	16.4	82.0	267	27	AV265878 AV265878
C 12	16.4	82.0	279	111	AW124239 UT-M-BH2
C 13	16.4	82.0	287	125	BB100422 BB100422
C 14	16.4	82.0	292	128	BB217229 BB217229
C 15	16.4	82.0	293	127	BB155980 BB155980
C 16	16.4	82.0	307	133	BB426512 BB426512
C 17	16.4	82.0	308	164	BE155459 PM1-HR035
C 18	16.4	82.0	343	238	AZ078760 RPCI-23-4
C 19	16.4	82.0	355	120	AW767597 d84b01.Y
C 20	16.4	82.0	366	114	AW311692 NAW055F1
C 21	16.4	82.0	372	8	AA517758 AA517758
C 22	16.4	82.0	376	23	A1698906 wa81g08.x
C 23	16.4	82.0	396	8	AA498831 V905603.r
C 24	16.4	82.0	406	15	AI030547 UT-R-CO-3
C 25	16.4	82.0	409	138	BE688775 BE688775
C 26	16.4	82.0	422	173	BE065373 H303D01-1
C 27	16.4	82.0	473	138	BE630423 ua20903.x
C 28	16.4	82.0	485	18	AI315397 u46b08.Y
C 29	16.4	82.0	486	4	AA254110 va08g07.r
C 30	16.4	82.0	508	18	AI314293 u124c05.x
C 31	16.4	82.0	509	174	BE141565 1a81b01.x
C 32	16.4	82.0	522	123	AW986420 u181c10.x
C 33	16.4	82.0	558	173	BE094146 ua72b07.x
C 34	16.4	82.0	566	139	BE753306 206275.MA
C 35	16.4	82.0	580	223	AQ001918 CIR-HSP-2
C 36	16.4	82.0	584	225	AQ159551 nxb00140
C 37	16.4	82.0	588	24	AI785259 u138c08.x
C 38	16.4	82.0	623	117	AW537380 G0115610
C 39	16.4	82.0	647	243	AZ433919 1M020J03
C 40	16.4	82.0	649	233	AQ780056 HS_3169_A
C 41	16.4	82.0	709	243	AZ409265 1M0180M1
C 42	16.4	82.0	849	170	BE866137 963066G12
C 43	16.4	82.0	1059	150	BE576271 602134137
C 44	16.4	82.0	1711	192	AK007734 MMS nmucu
C 45	16.4	80.0	458	225	AQ201343 RPCI11-43

## ALIGNMENTS

RESULT 1  
LOCUS AW982117/c  
DEFINITION SSS0027 Snaeda salsa ZAP cDNA library Snaeda maritima subsp. salsa  
CDNA similar to putative nematode-resistance protein, mRNA  
sequence.

ACCESSION AW982117  
VERSION AW982117.1 GI:8173690  
KEYWORDS EST.  
SOURCE Snaeda maritima subsp. salsa.  
ORGANISM Snaeda maritima subsp. salsa.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Caryophyllales; Caryophyllales; Chenopodiaceae; Snaeda.  
REFERENCE 1 (bases 1 to 405)  
Zhang, L., Ma, C.L., Wang, P.P., Sun, Y.F., Zhao, Y.X. and Zhang, H.  
Expressed sequence tags from a halophyte Snaeda salsa cDNA library

## JOURNAL COMMENT

Unpublished (2000)  
Contact: Hui Zhang  
Key Laboratory of Plant Stress Research  
The Biology Department of Shandong Normal University  
No. 88, Wenhua East Road, Jinan, Shandong Province, 250014, PRC  
Tel: (86)531-2960864  
Fax: (86)531-2966954  
Email: zhanghs@sdnu.edu.cn.

## FEATURES

## source

Location/Qualifiers

1..405  
/organism="Snaeda maritima subsp. salsa"  
/db\_xref="taxon:126914"  
/clone\_lib="Snaeda salsa ZAP cDNA library"

/dev\_stage="seeding"  
/note="Organ: aerial part tissue; Vector: lambda zap;  
Site\_1: EcoRI; Site\_2: XhoI; total RNA extraction from  
NaCl(400mM) treated Snaeda salsa by RNaseG kit(Promega);  
mRNA isolation by MESSAGEMAKER kit(GIBCO BRL);  
directional cDNA synthesis(EcoRI XhoI) by cDNA synthesis  
kit(STRATAGENE); the ZAP express library by GigapackIII  
Gold Cloning kit(STRATAGENE)"

## BASE COUNT

112 a 120 c 49 g 123 t 1 others

## ORIGIN

Query Match 84.0%; Score 16.8; DB 123; Length 405;  
Best Local Similarity 90.0%; Pred. No. 6; 5e-02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

## QY

1 gagtaggaagatagaacg 20  
||||||| |||||  
Db 219 GAGTAGGAGGAGTAGAACG 200

## RESULT 2

LOCUS BE859183/c  
DEFINITION SSS0490 Snaeda salsa ZAP cDNA library Snaeda maritima subsp. salsa  
CDNA similar to putative nematode-resistance protein, mRNA  
sequence.

## ACCESSION

BE859183  
VERSION BE859183.1 GI:10374851

## KEYWORDS

## ORGANISM

Snaeda maritima subsp. salsa.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Caryophyllales; Caryophyllales; Chenopodiaceae; Snaeda.

## REFERENCE

1 (bases 1 to 411)  
Zhang, L., Ma, C.L., Wang, P.P., Sun, Y.F., Zhao, Y.X. and Zhang, H.  
Expressed sequence tags from a halophyte Snaeda salsa cDNA library  
Unpublished (2000)  
Contact: Hui Zhang  
Key Laboratory of Plant Stress Research  
The Biology Department of Shandong Normal University  
No. 88, Wenhua East Road, Jinan, Shandong Province, 250014, PRC  
Tel: (86)531-2960864  
Fax: (86)531-2966954  
Email: zhanghs@sdnu.edu.cn.

## JOURNAL COMMENT

## FEATURES

## source

Location/Qualifiers

1..411  
/organism="Snaeda maritima subsp. salsa"  
/db\_xref="taxon:126914"  
/clone\_lib="Snaeda salsa ZAP cDNA library"

/dev\_stage="seeding"  
/note="Organ: aerial part tissue; Vector: lambda zap;  
Site\_1: EcoRI; Site\_2: XhoI; total RNA extraction from  
NaCl(400mM) treated Snaeda salsa by RNaseG kit(Promega);  
mRNA isolation by MESSAGEMAKER kit(GIBCO BRL);  
directional cDNA synthesis(EcoRI XhoI) by cDNA synthesis  
kit(STRATAGENE); the ZAP express library by GigapackIII  
Gold Cloning kit(STRATAGENE)"

## BASE COUNT

112 a 123 c 50 g 126 t

ORIGIN

Query Match 84.0% Score 16.8; DB 141; Length 411;  
Best Local Similarity 90.0%; Pred. No. 6.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gagtagaaggatagaacg 20  
|||||  
Db 219 GAGTAGAGGCGTAGAAGC 200

RESULT 3  
BE810166/c 511 bp mRNA EST 16-OCT-2000  
LOCUS SAS0388 Snaeda salsa ZAP cDNA library Snaeda maritima subsp. salsa  
DEFINITION cDNA similar to putative Hsipro-1 homolog, mRNA sequence.  
ACCESSION BE810166  
VERSION BE810166.1 GI:10241278  
KEYWORDS EST.  
SOURCE Snaeda maritima subsp. salsa.  
ORGANISM Snaeda maritima subsp. salsa.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Caryophyllales; Caryophyllales; Chenopodiaceae; Snaeda.  
REFERENCE 1 (bases 1 to 511)  
AUTHORS Zhang, L., Ma, C. L., Wang, P. P., Sun, Y. F., Zhao, Y. X. and Zhang, H.  
TITLE Expressed sequence tags from a halophyte Snaeda salsa cDNA library  
JOURNAL Unpublished (2000)  
COMMENT Contact: Hui Zhang  
Key Laboratory of Plant Stress Research  
The Biology Department of Shandong Normal University  
No. 88, Wenhua East Road, Jinan, Shandong Province, 250014, PRC  
Tel: (86)531-2960864  
Fax: (86)531-2966954  
Email: zhangh@snnu.edu.cn  
Location/Qualifiers

FEATURES  
source  
1. 511  
/organism="Snaeda maritima subsp. salsa"  
/db\_xref="taxon:126914"  
/clone\_lib="Snaeda salsa ZAP cDNA library"  
/dev\_strage="seedling"  
/note="Organ: aerial part tissue; Vector: lambda zap;  
Site\_1: EcoRI; Site\_2: XhoI; total RNA extraction from  
NaCl(400mM) treated Snaeda salsa by RNagent kit(Promega);  
mRNA isolation by MESSAGEMAKER kit(GIBCO BRL);  
directional cDNA synthesis(EcoRI XhoI) by cDNA synthesis  
kit(STRATAGEN); the ZAP express library by GigaPackIII  
Gold Cloning Kit(STRATAGENE)"

BASE COUNT 136 a 143 c 79 g 153 t

ORIGIN

Query Match 84.0% Score 16.8; DB 140; Length 511;  
Best Local Similarity 90.0%; Pred. No. 6.8e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gagtagaaggatagaacg 20  
|||||  
Db 219 GAGTAGAGGCGTAGAAGC 200

RESULT 4  
AZ609094/c 629 bp DNA GSS 13-DEC-2000  
LOCUS 1M043311R Mouse 10kb plasmid UUCG1M library Mus musculus genomic  
DEFINITION c10ne UUCG1M043311 R, DNA sequence.  
ACCESSION AZ609094  
VERSION AZ609094.1 GI:11731284  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS  
1 (bases 1 to 629)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Petersen, T., Kelly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
and Wright, D., Weiss, R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0433 row: 1, column: 11  
Seq primer: CACACAGGAACACACTATGAC  
Class: plasmid ends  
High quality sequence stop: 629.  
Location/Qualifiers

FEATURES  
source  
1. 629  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="UUCG1M043311"  
/clone\_lib="Mouse 10kb plasmid UUCG1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: pMD42uv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pMD42 (g114732114/gb1AF129072.1) a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 187 a 145 c 103 g 194 t

ORIGIN

Query Match 84.0% Score 16.8; DB 246; Length 629;  
Best Local Similarity 90.0%; Pred. No. 7.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gagtagaaggatagaacg 20  
|||||  
Db 203 GAGCAGGAAGGTACAAATG 184

RESULT 5  
AJ395794/c 814 bp mRNA EST 25-JAN-2001  
LOCUS AJ395794 dkfz426 Gallus gallus cDNA clone 24b16r1, mRNA sequence.  
DEFINITION AJ395794  
ACCESSION AJ395794  
VERSION AJ395794.1 GI:7127074  
KEYWORDS EST.  
SOURCE chicken.  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



REFERENCE	Archaeosauia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
AUTHORS	1 (bases 1 to 814) Abdrakhmanov,I., Lodgjin,D., Gerolt,P., Arakawa,H., Law,A., Plachy,J., Korn,B. and Buerstedde,J.M. A large database of chickenursal ESTS as a resource for the analysis of vertebrate gene function Genome Res. 10 (12), 2062-2069 (2000)
JOURNAL	20568495
MEDLINE	Contact: Buerstedde JM Cellular Immunology Heinrich-Pette-Institute Martinistr. 52, 20251 Hamburg, Germany Email: URL: //genetics.hpi.uni-hamburg.de/dt40est.html
COMMENT	Location/Qualifiers
FEATURES	Source 1..814 /organism="Gallus gallus" /strain="CB" /db_xref="taxon:9031" /clone="24b16r1" /clone_1lb="dkfz426" /tissue_type="Bursa Of Fabricius" Base Count 227 a 139 c 133 g 315 t Origin
Query Match	84.0%; Score 16.8; DB 104; Length 814; Best Local Similarity 90.0%; Pred. No. 7.4e+02; Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY	1 gagtaggaagatagaacac 20   Db 574 GTGTAGGAAGATTGAAAGC 555
RESULT	6
LOCUS	BE448549/c
DEFINITION	BE448549 196 bp mRNA EST 25-JUL-2000 uc92d01.v1 Soares:mammary_gland_MMLMG Mus musculus cDNA clone IMAGE:3369889 5', mRNA sequence.
ACCESSION	BE448549
VERSION	BE448549.1 GI:9448126
KEYWORDS	EST.
SOURCE	house mouse. Mus musculus
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgaps@mail.nih.gov THIS clone is available royalty-free through LNL ; contact the IMAGE Consortium (infoimage.lnl.gov) for further information. NCI:1079493 Seq primer: -40RP from Glbco High quality sequence stop: 159. Location/Qualifiers 1..196
FEATURES	Source /organism="Mus musculus" /db_xref="taxon:10090" /clone="IMAGE:3369889" /clone_1lb="Soares:mammary_gland_MMLMG" /sex="female(lactating)" /tissue_type="mammary gland" /lab_host="DH10B" /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not

BASE COUNT	40 a	39 c	34 g	83 t
ORIGIN				
Query Match	82.0%;	Score 16.4;	DB 167;	Length 196;
Best Local Similarity	94.4%;	Pred. No. 8.9e+02;		
Matches 17; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;
OY	2	agtaggaagatagaac	19	
Db	179	AGTAGGAAGATAGCAAC	162	
RESULT 7				
LOCUS	AW311691	207 bp	mRNA	EST
DEFINITION	XXNV055F11F Nsf Xylem Normal wood Vertical Pinus taeda cDNA clone			21-JAN-2000
ACCESSION	XXNV055F11	5', mRNA sequence.		
VERSION	AW311691			
KEYWORDS	AW311691.1	GI:6727337		
SOURCE	EST.			
ORGANISM	loblolly pine.			
REFERENCE	Pinus taeda			
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
TITLE	Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.			
JOURNAL	1 (bases 1 to 207)			
COMMENT	Sederoff, R.			
	Molecular Basis of Wood Formation in the Pine Megagenome			
	Unpublished (2000)			
	Contact: Johnson, Arthur			
	North Carolina State University			
	Tel: 919 515 7800			
	Fax: 919 515 7801			
	Email: ajohnson@ncsu.edu			
	Seq primer: T3.			
FEATURES				
source	Location/Qualifiers			
	1..207			
	/organism="Pinus taeda"			
	/db_xref="taxon:3352"			
	/clone="XXNV055F11"			
	/clone_11b="Nsf Xylem Normal wood Vertical"			
	/note="Vector: Bluescript SK; Site_1: Eco RI; The			
	sequences contain a 'cDNA adapter' between the EcoRI site			
	and the start of the EST. The adapter sequence is			
	'AATTGGCAGAG' "			
BASE COUNT	68 a	21 c	57 g	58 t
ORIGIN				3 others
Query Match	82.0%;	Score 16.4;	DB 114;	Length 207;
Best Local Similarity	94.4%;	Pred. No. 9e+02;		
Matches 17; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;
OY	1	gaatagaagagatagaac	18	
Db	24	GAGTAGGAAGATAGAAA	41	
RESULT 8				
LOCUS	BE648384	231 bp	mRNA	EST
DEFINITION	UI-M-BH2.1-aph-e-12-0-UI r1 NTR BMAP M.S3.1 Mus musculus cDNA clone			06-SEP-2000
ACCESSION	UI-M-BH2.1-aph-e-12-0-UI 5', mRNA sequence.			
VERSION	BE648384			
KEYWORDS	BE648384.1	GI:9974205		
SOURCE	EST.			
ORGANISM	house mouse.			
	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			









```

/lab host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCACAGGCTCTTTTCTTTTCTTTTCTTAA 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCCTGGATTAATTAATTCCTCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified Bluescript KS(+) after bulk excision from Lambda
FLC I."
```

Query Match	82.0%	Score 16.4	DB 128	Length 292
Best Local Similarity	94.4%	Pred. No. 9.6e+02		
Matches 17, Conservative	0	Mismatches 1	Indels 0	Gaps 0
QY	1	gaatagaagaagatagaa	18	
Db	182	GAGAAGGAAGATAGAAA	165	

RESULT	15
BI155980/c	
LOCUS	293 bp mRNA EST 29-JUN-2000
DEFINITION	BI155980 RIKEN full-length enriched, 16 days neonate thymus Mus musculus cDNA clone A130028k22 3', mRNA sequence.

VERSION	BB155980.1	GI:8811910
KEYWORDS	EST.	
SOURCE	house mouse.	

ORGANISM Mus musculus

REFERENCE  
AUTHORS  
Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;  
1 (bases 1 to 293)  
Konno, H., Aizawa, K., Akhita, S., Akiyama, J., Arakawa, T., Carninci

**TITLE**  
**JOURNAL**  
**COMMENT**

COMMENT

Contact: Yoshihide Hayashizaki  
Genome Exploration Research Group, Life Science Tsukuba Center,  
Genome Science Laboratory  
The Institute of Physical and Chemical Research (RIKEN), Genomic  
Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: +81-298-36-9013  
Fax: +81-298-36-9098  
Email: genome-res@rc.riken.go.jp,  
URL: <http://genome.irc.riken.go.jp/>  
Carnelli, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sakaki,  
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermostabilization and thermoactivation of thermolabile enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)

FEATURES  
source  
1. .293  
Location/Qualifiers

```

BASE COUNT
ORIGIN
65 a      82 c      41 g      104 t      1 others
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A130028M22"
/clone_lib="RIKEN full-length enriched, 16 days neonate
thymus"
/tissue_type="thymus"
/dev_stage="16 days neonate"
/lab_host="DH10B"
/notes="Site_1: Salt; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGGAGAGAGAGATCCAGAGGCTCTTTTCTTTTCTTTTCTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 185.0. Second
strand cDNA was prepared with the primer adaptor of
sequence [5' GAGGAGAGAGATTCGAGTTAATTAATTCCTCCCTCCCTCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified Bluescript KS(+) after bulk excision from Lambda
FLC I "
```

	Query Match	82.0%	Score 16.4	DB 127	Length 233
	Best Local Similarity	94.4%	Pred. No. 9.6e+02		
	Matches 17	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	3	gttaggaagatagaacg	20		
db	108	GAAGCAAGATGGAACG	91		

Search completed: October 16, 2001, 18:30:43  
Job time: 2476 sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2001, 19:00:58 ; Search time 88.04 Seconds  
(without alignments)  
43.006 Million cell updates/sec

Title: US-09-675-650-4

Sequence: 1 gagtagagagagatagaacg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing First 45 summaries

Database :

Issued\_Patents.NA:\*  
1: /cgn2\_6/ptodata/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/lna/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/lna/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/lna/PTOS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/lna/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	14.8	74.0	67	2	US-08-343-443B-14	Sequence 14, Appl
2	14.8	74.0	750	1	US-08-224-195-1	Sequence 1, Appl
3	14.8	74.0	780	1	US-08-224-195-2	Sequence 2, Appl
4	14.8	74.0	833	2	US-08-343-443B-97	Sequence 97, Appl
5	14.8	74.0	1192	1	US-08-340-539A-3	Sequence 3, Appl
6	14.8	74.0	1192	2	US-08-461-592B-3	Sequence 3, Appl
7	14.8	74.0	1456	3	US-09-308-406-1	Sequence 1, Appl
8	14.8	74.0	1930	1	US-07-982-112-1	Sequence 1, Appl
9	14.8	74.0	2371	2	US-08-343-443B-1	Sequence 1, Appl
10	14.8	74.0	2412	1	US-08-437-027-18	Sequence 18, Appl
11	14.8	74.0	246240	2	US-08-724-394A-20	Sequence 20, Appl
12	14.8	74.0	246240	2	US-08-724-394A-21	Sequence 21, Appl
13	14.8	74.0	246240	2	US-08-724-394A-22	Sequence 22, Appl
14	14.2	71.0	468	4	US-08-821-994-36	Sequence 36, Appl
15	14.2	71.0	507	4	US-08-821-994-75	Sequence 75, Appl
16	14.2	71.0	563	4	US-09-276-531-38	Sequence 38, Appl
17	14.2	71.0	743	4	US-08-821-994-37	Sequence 37, Appl
18	14.2	71.0	743	4	US-08-821-994-38	Sequence 38, Appl
19	14.2	71.0	926	4	US-08-106-507-1	Sequence 1, Appl
20	14.2	71.0	1102	4	US-08-821-994-86	Sequence 86, Appl
21	14.2	71.0	1390	4	US-08-821-994-61	Sequence 61, Appl
22	14.2	71.0	1434	4	US-08-821-994-62	Sequence 62, Appl
23	14.2	71.0	1441	4	US-08-821-994-63	Sequence 63, Appl
24	14.2	71.0	1661	4	US-08-821-994-82	Sequence 82, Appl
25	14.2	71.0	1674	4	US-08-096-762-199	Sequence 199, Appl
26	14.2	71.0	1674	4	US-09-042-353-59	Sequence 59, Appl
27	14.2	71.0	1728	3	US-08-379-802-1	Sequence 1, Appl

28	14.2	71.0	1728	3	US-09-048-129-1	Sequence 1, Appl
29	14.2	71.0	1728	4	US-09-048-079-1	Sequence 1, Appl
30	14.2	71.0	2334	1	US-08-062-632-4	Sequence 4, Appl
31	14.2	71.0	2407	2	US-08-895-522-2	Sequence 2, Appl
32	14.2	71.0	2407	3	US-09-195-391-2	Sequence 2, Appl
33	13.8	69.0	678	5	PCR-US94-04361-24	Sequence 24, Appl
34	13.8	69.0	1121	4	US-09-068-569-1	Sequence 1, Appl
35	13.8	69.0	1718	1	US-07-687-466-1	Sequence 1, Appl
36	13.8	69.0	1718	1	US-08-434-702-1	Sequence 1, Appl
37	13.8	69.0	1718	1	US-08-271-883-1	Sequence 1, Appl
38	13.8	69.0	1718	6	5168064-3	Sequence 1, Appl
39	13.8	69.0	2290	2	US-08-933-821-1	Sequence 1, Appl
40	13.8	69.0	2290	3	US-08-960-507-1	Sequence 1, Appl
41	13.8	69.0	2497	1	US-08-643-219-12	Sequence 12, Appl
42	13.8	69.0	2497	2	US-09-131-995-12	Sequence 12, Appl
43	13.8	69.0	2497	2	US-08-832-087B-12	Sequence 12, Appl
44	13.8	69.0	2497	3	US-08-851-350-12	Sequence 12, Appl
45	13.8	69.0	2497	4	US-09-132-154-12	Sequence 12, Appl

#### ALIGNMENTS

RESULT 1  
US-08-343-443B-14  
Sequence 14, Application US/08343443B  
Patent No. 5968734  
GENERAL INFORMATION:  
APPLICANT: Aurias, Alain  
APPLICANT: Delattre, Olivier  
APPLICANT: Desmaze, Chantal  
APPLICANT: Meiot, Thomas  
APPLICANT: Peter, Martine  
APPLICANT: Ploougaestel, Beatrice  
APPLICANT: Thomas, Gilles  
APPLICANT: Zucman, Jessica  
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAIID  
NUMBER OF SEQUENCES: 129  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weiser & Associates  
STREET: 230 South Filenehth Street  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: AEDIT 1.0 DOS text editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,443B  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR93/00494  
FILING DATE: 19-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/06123  
FILING DATE: 20-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 989,6121P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:

LENGTH: 67 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-343-443B-14

Query Match 74.0%; Score 14.8; DB 2; Length 67;  
Best Local Similarity 88.9%; Pred. No. 57;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gagtaggaagatagaaa 18  
||| ||||| |||||  
DB 7 GAGGAGGAAGAGAGAAA 24

RESULT 2  
US-08-224-195-1/C  
Sequence 1, Application US/08224195  
Patent No. 5496705  
GENERAL INFORMATION:  
APPLICANT: SUGANO, MITSUKO  
TITLE OF INVENTION: MONOCLONAL ANTIBODY  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
P.C.  
STREET: 1755 Jefferson Davis Highway, Fourth Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/224.195  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/929.377  
FILING DATE: 14-AUG-1992  
APPLICATION NUMBER: JP 3-229756  
FILING DATE: 16-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5496705man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 39-4161-0S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)521-4500  
TELEFAX: (703)486-2347  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 750 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Rattus norvegicus  
US-08-224-195-1

Query Match 74.0%; Score 14.8; DB 1; Length 750;  
Best Local Similarity 88.9%; Pred. No. 78;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 1 gagtaggaagatagaaa 18  
||| ||||| |||||

DB 584 GAGGAGGAAGAGAGAAA 567

RESULT 3  
US-08-224-195-2/C  
Sequence 2, Application US/08224195  
Patent No. 5496705

GENERAL INFORMATION:  
APPLICANT: SUGANO, MITSUKO  
TITLE OF INVENTION: MONOCLONAL ANTIBODY  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
P.C.

ADDRESSEE: P.C.  
STREET: 1755 Jefferson Davis Highway, Fourth Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/224.195  
FILING DATE:

CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/929.377  
FILING DATE: 14-AUG-1992  
APPLICATION NUMBER: JP 3-229756  
FILING DATE: 16-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5496705man F.

REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 39-4161-0S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)521-4500  
TELEFAX: (703)486-2347  
TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 780 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: rat  
US-08-224-195-2

Query Match 74.0%; Score 14.8; DB 1; Length 780;  
Best Local Similarity 88.9%; Pred. No. 79;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gagtaggaagatagaaa 18  
||| ||||| |||||  
DB 589 GAGGAGGAAGAGAGAAA 572

RESULT 4  
US-08-343-443B-97  
Sequence 97, Application US/0834343B  
Patent No. 5968734

GENERAL INFORMATION:  
APPLICANT: Aurias, Alain  
APPLICANT: Delattre, Olivier  
APPLICANT: Desmazes, Chantal  
APPLICANT: Melot, Thomas  
APPLICANT: Peter, Martine  
APPLICANT: Ploougastel, Beatrice

APPLICANT: Thomas, Gilles  
APPLICANT: Zucman, Jessica  
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
TITLE OF INVENTION: TRANSLOCATIONS  
NUMBER OF SEQUENCES: 129  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weiser & Associates  
STREET: 230 South Fifteenth Street  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: AEDIT 1.0 DOS text editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,443B  
CLASSIFICATION: 514  
FILING DATE: 18-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR93/00494  
FILING DATE: 19-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/06123  
FILING DATE: 20-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
TELECOMMUNICATION INFORMATION:  
REFERENCE/DOCKET NUMBER: 989.6121P  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 97:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 833 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-343-443B-97

Query Match 74.0%; Score 14.8; DB 2; Length 833;  
Best Local Similarity 88.9%; Pred. No. 79;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gagtaggaagataga 18  
||| ||||| |||||  
Db 575 GAGGAGGAAGAGAGAA 592

RESULT 5  
US-08-340-539A-3  
Sequence 3, Application US/08340539A  
Patent No. 5808025  
GENERAL INFORMATION:  
APPLICANT: Tedder, Thomas F.  
APPLICANT: Kansas, Geoffrey S.  
TITLE OF INVENTION: CHIMERIC SELECTINS AS SIMULTANEOUS  
TITLE OF INVENTION: BLOCKING AGENTS FOR COMPONENT SELECTIN FUNCTION  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/340,539A  
FILING DATE: 16-NOV-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/008,459  
FILING DATE: 25-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Gunnison, Jane  
REGISTRATION NUMBER: 38,479  
REFERENCE/DOCKET NUMBER: CG-104 CON  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-596-9000  
TELEFAX: 212-596-9090  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1192 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-340-539A-3

Query Match 74.0%; Score 14.8; DB 1; Length 1192;  
Best Local Similarity 88.9%; Pred. No. 83;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 gtgagaagatagaacg 20  
| ||||| |||||  
Db 700 GAAGGAAGATGAGAAAG 717

RESULT 6  
US-08-461-592B-3  
Sequence 3, Application US/08461592B  
Patent No. 5834425  
GENERAL INFORMATION:  
APPLICANT: Tedder, Thomas F.  
APPLICANT: Kansas, Geoffrey S.  
TITLE OF INVENTION: CHIMERIC SELECTINS AS SIMULTANEOUS  
TITLE OF INVENTION: BLOCKING AGENTS FOR COMPONENT SELECTIN FUNCTION  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
STREET: Ten Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/461,592B  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/340,539  
FILING DATE: 16-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/008,459  
FILING DATE: 25-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: James F. Haley, Jr.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: CG-104

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
TELEX: 14-8367
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1192 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-461-592B-3

Query Match      74.0% Score 14.8; DB 2; Length 1192;
Best Local Similarity 88.9%; Pred. No. 83;
Matches 16; Conservatively 0; Mismatches 2; Indels 0; Gaps 0;

Qy      3 gtagaaagatagaacg 20
        | ||||| ||||| ||
Db      700 GAAGCAACGATGAAGC 717

RESULT          7
US-09-308-406-1
Sequence 1, Application US/09308406
Patent No. 6159696
GENERAL INFORMATION:
APPLICANT: Dijkema, Rein
APPLICANT: van den Wijngaard, Arthur
TITLE OF INVENTION: Method and materials for the screening of therapeutic
TITLE OF INVENTION: agents for the prevention and/or treatment of
FILE REFERENCE: I/96227 US
CURRENT APPLICATION NUMBER: US/09/308,406
CURRENT FILING DATE: 1999-06-21
EARLIER APPLICATION NUMBER: PCT/EP97/06668
EARLIER FILING DATE: 1997-11-20
EARLIER APPLICATION NUMBER: EP96203283.5
EARLIER FILING DATE: 1996-11-22
NUMBER OF SEQ ID NOS: 16
SOFTWARE: patentln Ver. 2.1
SEQ ID NO 1
LENGTH: 1456
TYPE: DNA
ORGANISM: human
US-09-308-406-1

Query Match      74.0% Score 14.8; DB 3; Length 1456;
Best Local Similarity 88.9%; Pred. No. 86;
Matches 16; Conservatively 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 gagtagaaggatagaac 18
        ||| ||||| |||||
Db     1122 gaggaaggagaagaac 1139

RESULT          8
US-07-982-112-1/c
Sequence 1, Application US/07982112
Patent No. 5346813
GENERAL INFORMATION:
APPLICANT: BODENMULLER, Heinz
APPLICANT: DESSAUER, Andreas
TITLE OF INVENTION: METHOD FOR THE DETECTION OF A SMALL
TITLE OF INVENTION: CELL LUNG CARCINOMA
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nikaïdo, Marmelstein, Murray & Oram
STREET: 655 Fifteenth Street N.W. Suite 330
```

```

City: Washington
State: D.C.
Country: U.S.A.
Zip: 20005-5701

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/982,112
FILING DATE: 19921125
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/617,102
FILING DATE: 21-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Chin, Monica F.
REGISTRATION NUMBER: 36,105
REFERENCE/DOCKET NUMBER: 564-1906
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)638-5000
TELEFAX: (202)638-4810
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1930 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: CDS
LOCATION: 10..933
US-07-982-112-1

Query Match 74.0% Score 14.8; DB 1; Length 1930;
Best Local Similarity 88.9%; Pred. No. 89;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gagtagaagagatagaaa 18
||| ||||| |||||
Db 1309 GAGGAGAGAGAGAGAAA 1292

RESULT 9
US-08-343-443B-1
Sequence 1, Application US/08343443B
Patent No. 5968734
GENERAL INFORMATION:
APPLICANT: Aurias, Alain
APPLICANT: Delattre, Olivier
APPLICANT: Desmaze, Chantal
APPLICANT: Melot, Thomas
APPLICANT: Peter, Marline
APPLICANT: Ploougastel, Beatrice
APPLICANT: Thomas, Gilles
APPLICANT: Zucman, Jessica
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
TITLE OF INVENTION: TRANSLOCATIONS
NUMBER OF SEQUENCES: 129
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weiser & Associates
STREET: 230 South Fifteenth Street
City: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: AEDIT 1.0 DOS text editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,443B  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR93/00494  
FILING DATE: 19-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/06123  
FILING DATE: 20-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 989.6121P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2371 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 25..1992  
US-08-343-443B-1

Query Match 74.0%; Score 14.8; DB 2; Length 2371;  
Best Local Similarity 88.9%; Pred. No. 91;

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gagtaggaagatagaa 18  
||| ||||| |||||  
DB 7 GAGGAGGAAGAGAGAA 24

RESULT 10  
US-08-437-027-18  
Sequence 18, Application US/08437027  
Patent No. 5670317  
GENERAL INFORMATION:  
APPLICANT: Landanyi, Marc  
TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC  
TITLE OF INVENTION: SMALL ROUND CELL TUMOR  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/437,027  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 46416/JPM/CCA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400

TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2412 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
US-08-437-027-18

Query Match 74.0%; Score 14.8; DB 1; Length 2412;  
Best Local Similarity 88.9%; Pred. No. 91;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gagtaggaagatagaa 18  
||| ||||| |||||  
DB 7 GAGGAGGAAGAGAGAA 24

RESULT 11  
US-08-724-394A-20  
Sequence 20, Application US/08724394A  
Patent No. 5872237  
GENERAL INFORMATION:  
APPLICANT: Feder, John N.  
APPLICANT: Kronmal, Gregory S.  
APPLICANT: Lauer, Peter M.  
APPLICANT: Ruddy, David A.  
APPLICANT: Thomas, Winston  
APPLICANT: Tsuchihashi, Zenta  
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1  
TITLE OF INVENTION: Sequences and Antibodies Thereo  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,394A  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitts, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 017957-000100  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note="HLA-H.COMTIG"  
US-08-724-394A-20

Query Match 74.0%; Score 14.8; DB 2; Length 246240;

Best Local Similarity 88.9%, Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 2 agtaggaagatagaac 19  
|||||  
Db 141872 AGTAGGAAGATTAATAC 141889

RESULT 12  
US-08-724-394A-21  
Sequence 21, Application US/08724394A  
Patent No. 5872237  
GENERAL INFORMATION:  
APPLICANT: Feder, John N.  
APPLICANT: Krommal, Gregory S.  
APPLICANT: Lauer, Peter M.  
APPLICANT: Ruddy, David A.  
APPLICANT: Thomas, Winston  
APPLICANT: Tsuchihashi, Zenta  
APPLICANT: Wolff, Roger K.  
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1  
TITLE OF INVENTION: Sequences and Antibodies Thereto  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,394A  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitts, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 017957-000100  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "HLA-H.COMTIG"  
US-08-724-394A-21

Query Match 74.0%, Score 14.8; DB 2; Length 246240;  
Best Local Similarity 88.9%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 2 agtaggaagatagaac 19  
|||||  
Db 141872 AGTAGGAAGATTAATAC 141889

RESULT 13  
US-08-724-394A-22  
Sequence 22, Application US/08724394A  
Patent No. 5872237

GENERAL INFORMATION:  
APPLICANT: Feder, John N.  
APPLICANT: Krommal, Gregory S.  
APPLICANT: Lauer, Peter M.  
APPLICANT: Ruddy, David A.  
APPLICANT: Thomas, Winston  
APPLICANT: Tsuchihashi, Zenta  
APPLICANT: Wolff, Roger K.  
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1  
TITLE OF INVENTION: Sequences and Antibodies Thereto  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,394A  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitts, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 017957-000100  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "HLA-H.COMTIG"  
US-08-724-394A-22

Query Match 74.0%, Score 14.8; DB 2; Length 246240;  
Best Local Similarity 88.9%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 2 agtaggaagatagaac 19  
|||||  
Db 141872 AGTAGGAAGATTAATAC 141889

RESULT 14  
US-08-821-994-36  
Sequence 36, Application US/08821994A  
Patent No. 6228643  
GENERAL INFORMATION:  
APPLICANT: Greenland, Andrew J  
APPLICANT: Thomas, Didier RP  
APPLICANT: Jepson, Ian  
TITLE OF INVENTION: Promoters  
FILE REFERENCE: PPD 50108  
CURRENT APPLICATION NUMBER: US/08/821,994A  
CURRENT FILING DATE: 1997-03-22  
EARLIER APPLICATION NUMBER: PCT/GB97/00729  
EARLIER FILING DATE: 1997-03-18  
EARLIER APPLICATION NUMBER: GB 9606062.9  
EARLIER FILING DATE: 1996-03-22

```

; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
; LENGTH: 468
; TYPE: DNA
; ORGANISM: Brassica napus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(29), 439, 441
; OTHER INFORMATION: n is unknown
US-08-821-994-36

```

```

Query Match      71.0%; Score 14.2; DB 4; Length 468;
Best Local Similarity 84.2%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      2 agtaggaagatagaacg 20
      1 ||||| ||| |||||
Db      391 attagatcggagagaacg 409

```

```

RESULT 15
US-08-821-994-75
; Sequence 75, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 75
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(5), 10, 13, 15, 16, 28, 150, 276, 285, 292, 312, 337
; OTHER INFORMATION: n is uncertain
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 355, 357, 447, 477, 495
; OTHER INFORMATION: n is uncertain
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Consensus
US-08-821-994-75

```

```

Query Match      71.0%; Score 14.2; DB 4; Length 507;
Best Local Similarity 84.2%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY      2 agtaggaagatagaacg 20
      1 ||||| ||| |||||
Db      394 attagatcggagagaacg 412

```

Search completed: October 16, 2001, 19:01:18  
Job time: 4096 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2001, 19:01:36 ; Search time 164.45 Seconds  
(without alignments)  
76.364 Million cell updates/sec

Title: US-09-675-650-4

Perfect score: 20  
Sequence: 1 gagtagaagatagaacac 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

N.Geneseq.0601.\*  
1: /SIDSB/gcgdata/geneseq/geneseqn/NA1980.DAT.\*  
2: /SIDSB/gcgdata/geneseq/geneseqn/NA1981.DAT.\*  
3: /SIDSB/gcgdata/geneseq/geneseqn/NA1982.DAT.\*  
4: /SIDSB/gcgdata/geneseq/geneseqn/NA1983.DAT.\*  
5: /SIDSB/gcgdata/geneseq/geneseqn/NA1984.DAT.\*  
6: /SIDSB/gcgdata/geneseq/geneseqn/NA1985.DAT.\*  
7: /SIDSB/gcgdata/geneseq/geneseqn/NA1986.DAT.\*  
8: /SIDSB/gcgdata/geneseq/geneseqn/NA1987.DAT.\*  
9: /SIDSB/gcgdata/geneseq/geneseqn/NA1988.DAT.\*  
10: /SIDSB/gcgdata/geneseq/geneseqn/NA1989.DAT.\*  
11: /SIDSB/gcgdata/geneseq/geneseqn/NA1990.DAT.\*  
12: /SIDSB/gcgdata/geneseq/geneseqn/NA1991.DAT.\*  
13: /SIDSB/gcgdata/geneseq/geneseqn/NA1992.DAT.\*  
14: /SIDSB/gcgdata/geneseq/geneseqn/NA1993.DAT.\*  
15: /SIDSB/gcgdata/geneseq/geneseqn/NA1994.DAT.\*  
16: /SIDSB/gcgdata/geneseq/geneseqn/NA1995.DAT.\*  
17: /SIDSB/gcgdata/geneseq/geneseqn/NA1996.DAT.\*  
18: /SIDSB/gcgdata/geneseq/geneseqn/NA1997.DAT.\*  
19: /SIDSB/gcgdata/geneseq/geneseqn/NA1998.DAT.\*  
20: /SIDSB/gcgdata/geneseq/geneseqn/NA1999.DAT.\*  
21: /SIDSB/gcgdata/geneseq/geneseqn/NA2000.DAT.\*  
22: /SIDSB/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	16.8	84.0	792	20	AAZ12053
C 2	16.8	84.0	6147	21	AAZ53858
C 3	16.4	82.0	347	21	AAZ19786
C 4	15.8	79.0	300	20	AAZ13970
C 5	15.8	79.0	1926	21	AAA08075
C 6	15.8	79.0	2720	22	AAZ02348
C 7	15.8	79.0	2844	21	AAZ02349
C 8	15.8	79.0	611590	21	AAZ23303
C 9	15.4	77.0	6761	20	AAZ20517
C 10	15.2	76.0	485	20	AAV89036
C 11	15.2	76.0	620	21	AAZ09941

C 12	15.2	76.0	1544	18	AAZ86839
C 13	15.2	76.0	3127	20	AAZ40171
C 14	15.2	76.0	6774	21	AAZ39174
C 15	15.2	76.0	144460	21	AAZ93815
C 16	15.2	76.0	1038602	20	AAZ01425
C 17	15	75.0	300	20	AAZ13594
C 18	15	75.0	300	20	AAZ13588
C 19	14.8	74.0	233	22	AAZ44494
C 20	14.8	74.0	393	21	AAZ07987
C 21	14.8	74.0	401	21	AAZ28961
C 22	14.8	74.0	454	21	AAZ01659
C 23	14.8	74.0	735	4	AAZ30161
C 24	14.8	74.0	750	14	AAZ03214
C 25	14.8	74.0	760	22	AAZ22877
C 26	14.8	74.0	763	4	AAZ0162
C 27	14.8	74.0	763	11	AAZ006498
C 28	14.8	74.0	763	19	AAZ33298
C 29	14.8	74.0	814	21	AAZ75898
C 30	14.8	74.0	833	14	AAZ050688
C 31	14.8	74.0	876	4	AAZ30108
C 32	14.8	74.0	876	11	AAZ006497
C 33	14.8	74.0	876	19	AAZ33297
C 34	14.8	74.0	1192	15	AAZ071007
C 35	14.8	74.0	1342	21	AAZ49714
C 36	14.8	74.0	1346	21	AAZ41864
C 37	14.8	74.0	1443	19	AAZ94208
C 38	14.8	74.0	1456	19	AAZ32783
C 39	14.8	74.0	1533	21	AAZ49698
C 40	14.8	74.0	1535	21	AAZ40036
C 41	14.8	74.0	1686	2	AAZ10014
C 42	14.8	74.0	2371	14	AAZ050643
C 43	14.8	74.0	2403	18	AAZ97870
C 44	14.8	74.0	2412	18	AAZ97869
C 45	14.8	74.0	2412	18	AAZ97869

## ALIGNMENTS

RESULT 1	AAZ12053/C	standard; DNA; 792 BP.
ID	AAZ12053	
AC	AAZ12053	
XX		
DT	08-OCT-1999	(first entry)
XX		
DE	Neisseria gonorrhoeae complete ORF31 sequence.	
XX		
KW	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;	
KW	treatment; Neisseria infection; meningitis; septicaemia; gonorrhea; ss.	
XX		
OS	Neisseria gonorrhoeae.	
XX		
PN	W09924578-A2.	
XX		
PD	20-MAY-1999.	
XX		
PF	09-OCT-1998;	98MO-IB01665.
XX		
PR	01-SEP-1998;	98GB-0019016.
PR	06-NOV-1997;	97GB-0023516.
PR	14-NOV-1997;	97GB-0024190.
PR	18-NOV-1997;	97GB-0024386.
PR	27-NOV-1997;	97GB-0025158.
PR	10-DEC-1997;	97GB-0026147.
PR	14-JAN-1998;	98GB-0000755.
XX		
PA	(CHIR-) CHIRON SPA.	
XX		
PI	Grandi G, Masiagnanl V, Pizza M, Rappuoli R, Scarlato V;	
XX		
DR	WPI, 1999-327407/27.	

cDNA encoding rat  
W09904265 Seq ID N  
Bone morphogenetic  
Olfactory receptor  
Complete genome se  
Human gene express  
Human gene express  
Mouse dextran sodi  
Human secreted pro  
Human secreted pro  
Human secreted pro  
Sequence encoding  
Rat synaptophysin  
Human prostate can  
Sequence of the B.  
IFN-alpha 76 gene  
Escherichia coli t  
Human ORF1453  
Human Ews gene pro  
Sequence encoding  
IFN-alpha 76 gene.  
DNA encoding Inter  
Human Interferon-a  
Ilym-1 gene exon I  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Human BMP-4 promot  
Arabidopsis thalia  
Arabidopsis thalia  
Sequence encoding  
Human Ews gene clo  
cDNA encoding a ch  
cDNA encoding a ch

DR P-PSDB: AAY38591.  
 XX Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for  
 PT diagnosis, treatment and prevention of infection  
 XX  
 PS Claim 9; Page 150-151; 524pp; English.  
 CC Nucleotide sequences AA21972-212358 represent open reading frames  
 CC (ORFs) of *Neisseria meningitidis* and *N. gonorrhoeae* which encode  
 CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their  
 CC fragments, their nucleic acids and antibodies are used for diagnosis,  
 CC prevention (as vaccines) or treatment of *Neisseria* infections,  
 CC such as meningitis, septicaemia and gonorrhea. Both organisms  
 CC are closely related. Fragments of the nucleic acids are useful  
 CC as hybridisation probes and antisense reagents.  
 XX  
 SQ Sequence 792 BP; 210 A; 184 C; 189 G; 209 T; 0 other;

Query Match 84.0%; Score 16.8; DB 20; Length 792;  
 Best Local Similarity 90.0%; Pred. No. 57;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gagtaggaagatagaacg 20  
 |||||  
 DB 148 GAGTAGGAATGAAAGAACG 129

RESULT 2  
 AA23858/c  
 ID AA23858 standard; DNA; 6147 BP.  
 XX  
 AC AA23858;  
 XX  
 DT 21-MAR-2000 (first entry)  
 XX  
 DE *Neisseria gonorrhoeae* ORF 564 partial DNA sequence SEQ ID NO:1665.  
 XX  
 KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;  
 KW antibacterial; gene therapy; ds.  
 XX  
 OS *Neisseria gonorrhoeae*.  
 OS  
 PN WO957280-A2.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 30-APR-1999; 99WO-US09346.  
 XX  
 PR 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 02-SEP-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX  
 DR WPI: 2000-062150/05.  
 DR P-PSDB: AAY75096.  
 XX  
 PT Novel *Neisseria* polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics  
 XX  
 PS Claim 7; Page 849-851; 1453pp; English.

XX  
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941  
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC *Neisseria* bacteria (e.g. meningitis and septicaemia), to detect the  
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.  
 XX  
 SQ Sequence 6147 BP; 1998 A; 1476 C; 1444 G; 1229 T; 0 other;

Query Match 84.0%; Score 16.8; DB 21; Length 6147;  
 Best Local Similarity 90.0%; Pred. No. 62;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gagtaggaagatagaacg 20  
 |||||  
 DB 148 GAGTAGGAATGAAAGAACG 129

RESULT 3  
 AAC19786  
 ID AAC19786 standard; cDNA; 347 BP.  
 XX  
 AC AAC19786;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein 5' EST, SEQ ID NO: 23861.  
 XX  
 DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.  
 XX  
 OS *Homo sapiens*.  
 XX  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PE 21-FEB-2000; 2000EP-0200610.  
 XX  
 PF 26-FEB-1999; 99US-0122487.  
 XX  
 PR (GEST ) GENSET.  
 XX  
 PA Dunas Mline Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI: 2000-500381/45.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 PS Claim 1; SEQ ID 23861; 71pp + CD-ROM; English.  
 XX  
 CC The present sequence is one of a large number of 5' ESTs derived from  
 CC cDNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.

XX Sequence 347 BP; 132 A; 51 C; 57 G; 105 T; 2 other;

Query Match  
Best Local Similarity 94.4%; Pred. No. 83;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 agtaggaagatagaac 19  
DB 246 agtaggaagatagaac 263

RESULT 4  
AA213970  
ID AA213970 standard; cDNA; 300 BP.

XX AA213970;  
XX 12-OCT-1999 (first entry)

DE Human gene expression product cDNA sequence SEQ ID NO:1439.

XX Human; gene; gene expression product; diagnosis; therapy; probe;  
XX detection; mapping; tissue typing; profiling; forensic; cancer;  
XX genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

OS Homo sapiens.

XX W0938972-A2.

PN 05-AUG-1999.

PD 28-JAN-1999; 99WO-US01619.

PF 03-APR-1998; 98US-0080666.

PR 28-JAN-1998; 98US-0072910.

PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.

PR 03-APR-1998; 98US-0080515.

XX (CHIR ) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

PI Escobedo J, Garcia PD, Garcia V, Gliese K, Innis MA;

PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

PI Stache-Crain B, Sudduth-Klinger J, Williams LT;

PI WPI; 1999-494092/41.

DR Novel human genes and their expression products which are

XX differentially expressed in different cell types

XX Claim 1; Page 987; 2479pp; English.

CC The present invention describes a library of human polynucleotides  
CC comprising the sequences given in AA212532 to AA217779. Also described is  
CC a method of detecting differentially expressed genes correlated with the  
CC cancerous state of a mammalian cell, comprising detecting at least one  
CC differentially expressed gene product in a test sample from a cell  
CC suspected of being cancerous, where the gene product is encoded by one  
CC of the 5248 polynucleotide sequences given in AA212532 to AA217779. The  
CC polynucleotides can be used as a source of primers and probes, which can  
CC be used for a variety of purpose, e.g. detection of expression levels,  
CC mapping, tissue typing or profiling, forensics, genetic analysis and  
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
CC can be used for raising antibodies for experimental, diagnostic and  
CC therapeutic purposes. The polynucleotides may also be used to construct

CC arrays for diagnostics (which may be used to determine function of an  
CC encoded protein); and to detect differences in expression levels between  
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
CC identify a genetic predisposition or susceptibility to a disease such as  
CC cancer). The polynucleotides of the invention are especially used in the  
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
CC and lung cancer. The polynucleotides can also be used to screen for  
CC peptide analogues and antagonists.

XX Sequence 300 BP; 105 A; 42 C; 58 G; 95 T; 0 other;

Query Match  
Best Local Similarity 79.0%; Score 15.8; DB 20; Length 300;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 agtaggaagatagaac 20  
DB 16 aataggagatagaac 34

RESULT 5  
AAA08075/c  
ID AAA08075 standard; cDNA to mRNA; 1926 BP.

XX AAA08075;  
XX 21-JUN-2000 (first entry)

DE Human NPw38BP1 transcription related protein encoding cDNA SEQ ID NO:2.

XX Human; NPw38BP1; NPw38; transcription related protein; diagnosis;  
XX genetic disease; gene therapy; antitumour; ds.

OS Homo sapiens.

XX Key CDS

FT 1.1926

FT /tag=

XX /product= "NPw38BP1 transcription related protein"

XX JP2000060562-A.

PD 29-FEB-2000.

PF 21-AUG-1998; 98JP-0235901.

PR 21-AUG-1998; 98JP-0235901.

PR (KAGA-) KAGAKU GIYUTSU SHINKO JIGYODAN.

PA (KATO/) KATO S.

PA WPI; 2000-249677/22.

DR P-PSDB; AAY82327.

DR Human transcription-related protein and a cDNA encoding said protein

XX useful for screening for new antitumor agents and for diagnosis

XX Claim 5; Page 8; 10pp; Japanese.

XX The present sequence encodes a human transcription-related protein.

XX The protein and an antibody against the protein, are useful for

XX diagnosis and treatment of various diseases. The human cDNA is useful

XX as a probe for genetic diagnosis and a gene source for gene therapy.

XX A new antitumor agent can be detected by screening a low molecular

XX compound combining with the protein.

XX Sequence 1926 BP; 546 A; 529 C; 439 G; 412 T; 0 other;

Query Match  
Best Local Similarity 79.0%; Score 15.8; DB 21; Length 1926;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gagtaggaagatagaac 19  
|| |||||  
Db 556 GAACAGCAGCATGAAAC 538

RESULT 6  
AAD02348/c  
ID AAD02348 standard; cDNA; 2720 BP.

AC AAD02348;  
XX  
DT 24-APR-2001 (first entry)

DE Human RNA metabolism protein (RMEP-5) cDNA.

KW Human: RNA metabolism protein; RMEP; noctropic; neuroleptic; antitumor;  
KW transquilliser; antianaemic; antidiabetic; immunosuppressive; cytosstatic;  
KW antisthmatic; antinflammatory; anti-HIV; human immunodeficiency virus;  
KW antiarthritic; antiatherosclerotic; antiatherosclerotic; antiallergic;  
KW antirheumatoid; antiparkinsonian; antihypoid; nephrotrophic; antigout;  
KW thyromimetic; RMEP expression modulator; transgenic; spinal cord disease;  
KW nervous system disorder; Alzheimer's disease; therapy; gene therapy;  
KW neuromuscular disorder; hepatitis; cancer; cell proliferative disorder;  
KW peripheral nervous system disorder; cirrhosis; cranial nerve disorder;  
KW progressive neural autonomic nervous system disorder; Addison's disease;  
KW amyotrophic lateral sclerosis; autoimmune disorder; drug screening; ss.

OS Homo sapiens.

XX  
FH Key Location/Qualifiers  
FT CDS 192..2117  
FT /tag= a  
FT /product= "Human RNA metabolism protein (RMEP-5)"

XX W0200078952-A2.  
XX  
XX 28-DEC-2000.  
XX  
XX 15-JUN-2000; 2000WO-US16644.  
XX  
XX 17-JUN-1999; 99US-0139922.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX (BAUG/) BAUGHN M R.  
XX  
XX Bandman O, Yue H, Lal P, Tang YT, Reddy R, Azimzal Y;  
XX  
XX WPI: 2001-102723/11.  
XX P-PSDB; NAY72165.

PT New human RNA metabolism proteins (RMEP), useful for diagnosing,  
PT treating, preventing nervous system, cell proliferative,  
PT autoimmune/inflammatory disorders associated with abnormal expression  
PT of RMEP -

PS Example 1; Page 97-98; 103pp; English.

XX  
XX The present sequence is human RNA metabolism protein (RMEP-5) cDNA  
XX (Clone ID 1858421) obtained from PROSNOR18 cDNA library.  
XX Agonists and antagonists of RMEP cDNA are useful for treating diseases or  
XX conditions associated with altered expression of functional RMEP. RMEP  
XX sequence or their mammalian homologues are useful for creating 'knock  
XX out' or 'knock in' humanised animals or transgenic animals to model  
XX human disease. RMEP sequence is useful in the diagnosis, prevention and  
XX treatment of nervous system disorders e.g. Alzheimer's disease, Pick's  
XX disease, Huntington's disease, Parkinson's disease, amyotrophic lateral  
XX sclerosis, and other motor neuron disorders, progressive neural autonomic  
XX nervous system disorders, cranial nerve disorders, spinal cord diseases,  
XX muscular dystrophy and other neuromuscular disorders, peripheral nervous  
XX system disorders, mental disorders including anxiety and schizophrenia,  
XX amnesia etc, cell proliferative disorders e.g. actinic keratosis,  
XX arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, mixed connective

CC tissue disease (MCTD), cancers e.g. adenocarcinoma, leukemia, lymphoma,  
CC melanoma etc., and autoimmune/inflammatory disorders such as acquired  
CC immuno deficiency syndrome (AIDS), Addison's disease, allergies, anaemia,  
CC asthma, diabetes mellitus, rheumatoid arthritis, Grave's disease and  
CC autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED).  
CC RMEP cDNA is useful for somatic or germ-line gene therapy. RMEP sequence  
CC is useful several drug screening assays.

XX  
XX  
XX Sequence 2720 BP; 781 A; 669 C; 604 G; 666 T; 0 other;

SQ

Query Match 79.0%; Score 15.8; DB 22; Length 2720;  
Best Local Similarity 89.5%; Pred. No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gagtaggaagatagaac 19  
|| |||||  
Db 747 GAACAGCAGCATGAAAC 729

RESULT 7  
AAC98219/c  
ID AAC98219 standard; cDNA; 2844 BP.

AC AAC98219;  
XX  
DT 09-MAR-2001 (first entry)

DE Human colon cancer antigen nucleotide sequence SEQ ID NO:229.

XX  
XX  
KW Human: colon cancer; colon cancer antigen; diagnosis; detection;  
KW identification; cytosstatic; cardioactive; neuroprotective; vulnerrary;  
KW immunomodulatory; muscular; gynaecological; gastrointestinal;  
KW nephrotrophic; antinfertive; antibacterial; gene therapy; wound;  
KW neural disorder; immune system disorder; muscular disorder;  
KW reproductive disorder; gastrointestinal disorder; renal disorder;  
KW infectious disease; cardiovascular disorder; ss.

XX  
XX Homo sapiens.  
XX  
XX W0200055351-A1.  
XX  
XX 21-SEP-2000.  
XX  
XX 08-MAR-2000; 2000WO-US05883.  
XX  
XX 12-MAR-1999; 99US-0124270.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Ruben SM;  
XX  
XX WPI: 2000-587534/55.  
XX P-PSDB; AAB53462.

PT Colon cancer associated gene sequences, referred to as colon cancer  
PT antigens, useful for the treatment, prevention, and diagnosis of colon  
PT disorders such as colon cancer -

PS Claim 1; Page 650-651; 2104pp; English.

XX  
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
XX called human colon cancer antigens, given in AAB53234 to AAB54006. The  
XX human colon cancer antigens can have cytosstatic, cardioactive, muscular;  
XX neuroprotective, immunomodulatory, gynaecological, gastrointestinal,  
XX vulnerrary, nephrotrophic, antinfertive and antibacterial activities, and  
XX can be used in gene therapy. The colon cancer antigen polynucleotides,  
XX proteins and antibodies to the proteins are useful for the prevention,  
XX treatment and diagnosis of colon disorders, such as colon cancer. The  
XX polynucleotides may be used in diagnostics and research, such as for  
XX chromosome identification, and as hybridisation probes. The proteins  
XX may also be used to prevent diseases such as neural disorders, immune  
XX system disorders, muscular disorders, reproductive disorders,

CC gastrointestinal disorders, wounds, renal disorders, infectious  
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and  
CC AAB54007 represent sequences used in the exemplification of the present  
CC invention.

SO Sequence 2844 BP; 816 A; 691 C; 625 G; 711 T; 1 other;

Query Match 79.0%; Score 15.8; DB 21; Length 2844;  
Best Local Similarity 89.5%; Pred. No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gagtagaagaagatagaac 19  
||| ||||| ||||| |||||  
DB 743 GAGAGGAGGAGATAGAAC 725

## RESULT 8

AAF22303/C  
ID AAF22303 standard; DNA; 611590 BP.

AC AAF22303;

DT 20-MAR-2001 (first entry)

DE Arabidopsis thaliana chromosome 2 centromere.

KM Centromere; michrosome; vector; ds.

OS Arabidopsis thaliana.

PN WO200055325-A2.

PD 21-SEP-2000.

PF 17-MAR-2000; 2000WO-US07392.

PR 18-MAR-1999; 99US-0125219.

PR 01-APR-1999; 99US-0127409.

PR 18-MAY-1999; 99US-0134770.

PR 13-SEP-1999; 99US-0153584.

PR 17-SEP-1999; 99US-0154603.

PA (UYCH-) UNIV CHICAGO.

PI Preuss D, Copenhaver G, Keith K;

DR WPI; 2000-587529/55.

PT Recombinant DNA construct comprising a plant centromere, useful for  
PT producing stably inherited michrosomes which can serve as vectors for  
PT the construction of transgenic plant and animal cells

PS Claim 45; Page 820-959; 1449pp; English.

CC The present invention relates to a recombinant DNA construct of a plant  
CC (Arabidopsis thaliana) centromere. The constructs are useful for  
CC producing stably inherited michrosomes which can serve as vectors for  
CC the construction of transgenic plant and animal cells expressing  
CC selected proteins such as hormones, enzymes, interleukins, clotting  
CC factors, cytokines, antibodies, and growth factors.

SO Sequence 611590 BP; 181893 A; 124460 C; 120254 G; 184983 T; 0 other;

Query Match 79.0%; Score 15.8; DB 21; Length 611590;  
Best Local Similarity 89.5%; Pred. No. 1.9e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gagtagaagaagatagaac 19  
||| ||||| ||||| |||||  
DB 476561 GAGAGGAGGAGATAGAAC 476543

## RESULT 9

AA20517  
ID AAX20517 standard; DNA; 6761 BP.

AC AAX20517;

DT 05-MAY-1999 (first entry)

DE Polynucleotide sequence from the genome of Treponema pallidum.

KM Treponema pallidum infection; syphilis; Borrelia infection; animal;  
KM enzyme production; ds.

OS Treponema pallidum.

PN WO9859034-A2.

PD 30-DEC-1998.

PF 23-JUN-1998; 98WO-US13041.

PR 24-JUN-1997; 97US-0050667.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Fraser CM;

DR WPI; 1999-081273/07.

PT New isolated Treponema pallidum nucleic acids - used to develop  
PT products for the detection, diagnosis, characterisation, prevention  
PT and therapy of T. pallidum infections, particularly syphilis

PS Claim 1; Page 285-289; 1150pp; English.

CC AAX20500-21243 represent polynucleotide sequences from the genome of  
CC Treponema pallidum. The sequences can be used for detection,  
CC diagnosis, characterisation, prevention and therapy for T. pallidum  
CC infections, particularly syphilis. They can also be used for detecting  
CC diseases related to Borrelia infections in animals, and for the  
CC production of biosynthetic products such as enzymes.

SO Sequence 6761 BP; 1693 A; 1585 C; 1749 G; 1723 T; 11 other;

Query Match 77.0%; Score 15.4; DB 20; Length 6761;  
Best Local Similarity 94.1%; Pred. No. 2.7e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gagtagaagaagatagaac 17  
||| ||||| ||||| |||||  
DB 292 gagtagaagaagatagaac 308

PS Claim 1; Page 285-289; 1150pp; English.

PT New isolated Treponema pallidum nucleic acids - used to develop  
PT products for the detection, diagnosis, characterisation, prevention  
PT and therapy of T. pallidum infections, particularly syphilis

PS Claim 1; Page 285-289; 1150pp; English.

CC AAX20500-21243 represent polynucleotide sequences from the genome of  
CC Treponema pallidum. The sequences can be used for detection,  
CC diagnosis, characterisation, prevention and therapy for T. pallidum  
CC infections, particularly syphilis. They can also be used for detecting  
CC diseases related to Borrelia infections in animals, and for the  
CC production of biosynthetic products such as enzymes.

SO Sequence 6761 BP; 1693 A; 1585 C; 1749 G; 1723 T; 11 other;

Query Match 77.0%; Score 15.4; DB 20; Length 6761;  
Best Local Similarity 94.1%; Pred. No. 2.7e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gagtagaagaagatagaac 17  
||| ||||| ||||| |||||  
DB 292 gagtagaagaagatagaac 308

PS Claim 1; Page 285-289; 1150pp; English.

PT New isolated Treponema pallidum nucleic acids - used to develop  
PT products for the detection, diagnosis, characterisation, prevention  
PT and therapy of T. pallidum infections, particularly syphilis

PS Claim 1; Page 285-289; 1150pp; English.

XX 15-OCT-1998.  
 PD 10-APR-1998; 98MO-US06956.  
 PF 10-APR-1997; 97US-0837312.  
 PR 10-APR-1997; 97US-0837312.  
 XX (GEMV ) GENETICS INST INC.  
 PA  
 PI Agostino MJ, Jacobs K, Lavallic ER, McCoy JM, Merberg D;  
 PI Racie LA, Spaulding V, Treacy M;  
 XX WPI; 1999-070078/06.  
 DR  
 XX New polynucleotides encoding human secreted proteins - derived from  
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
 PT ovary, pituitary, retina and colon cDNA libraries  
 XX  
 PS Claim 1; Page 595-596; 641pp; English.  
 XX  
 CC The present sequence represents an expressed sequence tag (EST), and is  
 CC a polynucleotide of the invention. The polynucleotides of the invention  
 CC are all secreted EST sequences isolated from a variety of human tissue  
 CC sources. The EST sequences and proteins encoded by them are predicted to  
 CC have useful biological activities which would make them suitable for  
 CC treating, preventing or ameliorating medical conditions in humans and  
 CC animals, although no supporting data is given. Suggested activities  
 CC include nutritional activity, immune stimulating or suppressing activity,  
 CC hematopoietic regulating activity, tissue growth activity, hemostatic  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, anti-inflammatory  
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC activity. The EST sequences are also stated to be useful for gene  
 CC therapy.  
 XX  
 SQ Sequence 485 BP; 204 A; 58 C; 117 G; 106 T; 0 other;

Query Match 76.0%; Score 15.2; DB 20; Length 485;  
 Best Local Similarity 85.0%; Pred. No. 2.9e+02;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gagtagaagagatagaacg 20  
 |||||  
 Db 440 gagtagaagagatagaacg 459

RESULT 11  
 AAF09941  
 ID AAF09941 standard; cDNA; 620 BP.  
 XX  
 AC AAF09941;  
 XX  
 DT 13-MAR-2001 (first entry)  
 XX  
 DE Fusarium venenatum EST SEQ ID NO:2464.  
 XX  
 KW Multiple gene expression; filamentous fungal cell; EST;  
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KW Aspergillus oryzae; Trichoderma reesei; Identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.  
 XX  
 OS Fusarium venenatum.  
 XX  
 PN WO200056762-A2.  
 XX  
 PD 28-SEP-2000.  
 XX  
 PF 22-MAR-2000; 2000WO-US07781.  
 XX  
 PR 22-MAR-1999; 99US-0273623.  
 XX

PA (NOVO ) NOVO NORDISK BIOTECH INC.  
 PA (NOVO ) NOVO NORDISK AS.  
 XX  
 XX Berka RM, Rey MM, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;  
 PI WPI; 2000-594572/56.  
 DR  
 XX Monitoring differential expression of genes in filamentous fungal cells  
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags -  
 XX  
 PS Claim 86; Page 1287; 3161pp; English.  
 XX  
 CC The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring  
 CC the global expression of genes from FF cells allows the production  
 CC potential of the microorganisms to be improved. New genes may be  
 CC discovered, possible functions of unknown open reading frames can be  
 CC identified and gene copy number variation and stability can be  
 CC monitored. The expression of genes can be used to study how FF cells  
 CC adapt to changes in culture conditions, environmental stress, spore  
 CC morphogenesis, recombination, metabolic or catabolic pathway  
 CC engineering. Using ESTs provides several advantages over genomic or  
 CC random cDNA clones including elimination of redundancy as one spot on an  
 CC array equals one gene or open reading frame, and organisation of the  
 CC microarrays based on function of the gene products to facilitate  
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus  
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and  
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
 CC all specifically claimed in the present invention.  
 XX  
 SQ Sequence 620 BP; 173 A; 110 C; 143 G; 189 T; 5 other;

Query Match 76.0%; Score 15.2; DB 21; Length 620;  
 Best Local Similarity 85.0%; Pred. No. 3e+02;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gagtagaagagatagaacg 20  
 |||||  
 Db 177 gagtagaagagatagaacg 196

RESULT 12  
 AAT86839/C  
 ID AAT86839 standard; cDNA; 1544 BP.  
 XX  
 AC AAT86839;  
 XX  
 DT 19-DEC-1997 (first entry)  
 XX  
 DE cDNA encoding rat CC chemokine receptor.  
 XX  
 KW rat; CC chemokine receptor; screen; binding; ligand; ds.  
 KW  
 XX Rattus rattus.  
 XX  
 OS  
 XX  
 FH Key location/Qualifiers  
 FT 33..1100  
 FT CDS /\*tag= a  
 XX  
 XX JP09227599-A.  
 XX  
 PD 02-SEP-1997.  
 XX  
 PF 22-FEB-1996; 96JP-0035192.  
 XX

XX 22-FEB-1996; 96JP-0035192.  
XX (TAKE ) TAKEDA CHEM IND LTD.  
XX WPI; 1997-486426/45.  
XX P-PSDB; AAM29179.  
XX CC Chemokine receptor protein - useful to screen for novel binding  
XX compounds  
XX Claim 4; Page 21-22; 26pp; Japanese.  
XX This cDNA encodes a rat CC chemokine receptor. The receptor can be used  
XX to screen for novel binding compounds and for preparation of antibodies  
XX or antiserum.  
XX Sequence 1544 BP; 415 A; 347 C; 330 G; 452 T; 0 other;  
XX  
XX Query Match 76.0%; Score 15.2; DB 18; Length 1544;  
XX Best Local Similarity 85.0%; Pred. No. 3.1e+02;  
XX Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Oy 1 gagtaggaagatagaacg 20  
DB 785 GAGGAGGAAGATTAGACGC 766  
||| ||||| ||||| ||  
RESULT 13  
AAZ40171  
ID AAZ40171 standard; DNA; 3127 BP.  
XX AAZ40171;  
XX 02-JUL-1999 (first entry)  
XX WO9904265 Seq ID No: 683.  
XX Cancer associated antigen; diagnosis; research; treatment; human;  
XX breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
XX prostate cancer; ss.  
XX Homo sapiens.  
XX WO9904265-A2.  
XX 28-JAN-1999.  
XX 15-JUL-1998; 98WO-US14679.  
XX 22-JUN-1998; 98US-0102322.  
XX 17-JUL-1997; 97US-0896164.  
XX 10-OCT-1997; 97US-0061599.  
XX 10-OCT-1997; 97US-0061765.  
XX 10-OCT-1997; 97US-0948705.  
XX 11-OCT-1997; 97GB-0021697.  
XX (LUDW-) LUDWIG INST CANCER RES.  
XX Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;  
XX Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;  
XX Tureci O;  
XX WPI; 1999-132448/11.  
XX New isolated cancer associated nucleic acids and polypeptides -  
XX isolated using sera from cancer patients, used to develop products  
XX for the diagnosis, monitoring or treatment of cancers  
XX Claim 67; Page 725; 787pp; English.  
XX The invention relates to a method for diagnosing a disorder characterised

CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.  
XX Sequence 3127 BP; 1164 A; 500 C; 748 G; 705 T; 10 other;  
XX  
XX Query Match 76.0%; Score 15.2; DB 20; Length 3127;  
XX Best Local Similarity 85.0%; Pred. No. 3.2e+02;  
XX Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Oy 1 gagtaggaagatagaacg 20  
DB 1002 gagtaggaagatgagagag 1021  
||| ||||| ||||| ||  
RESULT 14  
AAZ39174  
ID AAZ39174 standard; DNA; 6774 BP.  
XX AAZ39174;  
XX 06-MAR-2000 (first entry)  
XX Bone morphogenetic protein 4 5' upstream gene sequence with exon 1-3.  
XX DE Bone morphogenetic protein 4; BMP-4; promoter; cartilage;  
XX KW bone disease; osteohyperplasia; osteometastasis; orthopaedic surgery;  
XX KW osteoarthritis; arthroplastis; tumour dissection; bone reconstruction;  
XX KW spinal fusion; vertebral canal enlargement; congenital cartilage disease;  
XX KW dysosteogenesis; achondroplasia; palatoschisis; mandible reconstruction;  
XX KW residual ridge construction; osteoporosis; morphogenesis; hyperplasia;  
XX ss.  
XX Homo sapiens.  
XX WO9957145-A2.  
XX 11-NOV-1999.  
XX 22-APR-1999; 99WO-IB00732.  
XX 30-APR-1998; 98JP-0120173.  
XX (HMRI ) HOECHST MARION ROUSSEL LTD.  
XX Kawai S, Sugiyura T;  
XX WPI; 2000-062142/05.  
XX Human bone morphogenetic protein 4 (BMP-4) promoter used to identify  
XX low molecular weight compounds which regulate the expression of BMP-4  
XX Claim 1; Page 19-22; 23pp; English.  
XX The present sequence represents the 5' upstream gene sequence including  
XX the exon 1 through exon 3 regions of human bone morphogenetic protein 4  
XX (BMP-4). This region of DNA and a reporter gene can be introduced into  
XX a host cell, and used to identify low molecular weight compounds which  
XX regulate the expression of human BMP-4. The low molecular weight  
XX compounds and their derivatives have morphogenetic activity and  
XX inhibiting activity for bone and cartilage through the expression of  
XX human BMP-4, and are useful as preventative or therapeutic agents for





Query Match 76.0%; Score 15.2; DB 21; Length 144460;  
Best Local Similarity 85.0%; Pred. No. 3.6e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Oy 1 gaatagaagaatagaacg 20  
||| ||||| |||||  
Db 127449 GAGGAGAGAGAGAGAGG 127430

Search completed: October 16, 2001, 19:01:56  
Job time: 3919 sec

**THIS PAGE BLANK (USPTO)**



Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 1 MFHISPPKYPHIOEAKKA 21  
:|:::|:| | |:::|  
DB 1212 VFINVRTPKYKHHYQCCEDHA 1232

RESULT 2

ID 09WJDS PRELIMINARY; PRT; 2485 AA.  
AC 09WJDS;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE NONSTRUCTURAL POLYPROTEIN.  
OS Venezuelan equine encephalitis virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
OC Alphavirus.  
OX NCBI\_TaxID=11036;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ZPC738;  
RX MEDLINE=99214366; PubMed=10196323;  
RA Wang E., Barrera R., Boshell J., Ferro C., Freter J.E., Navarro J.C.,  
RA Salas R., Vasquez C., Weaver S.C.;  
RT "Genetic and phenotypic changes accompanying the emergence of  
RT enzootic subtype IC Venezuelan equine encephalitis viruses from an  
RT enzootic subtype ID progenitor";  
RL J. Virol. 73:4266-4271(1999).  
DR EMBL; AF100566; AAD27802.1; -.  
DR InterPro; IPR000606; -.  
DR InterPro; IPR001788; -.  
DR InterPro; IPR002589; -.  
DR InterPro; IPR003620; -.  
DR Pfam; PF00978; RNA\_dep\_RNAPol2; 2.  
DR Pfam; PF01443; Viral\_helicase1; 1.  
DR Pfam; PF01661; DUF27; 1.  
DR Pfam; PF01707; Peptidase\_C9; 1.  
KW Polyprotein.  
SQ SEQUENCE 2485 AA; 276784 MW; 0390819206595822 CRC64;

Query Match 41.5%; Score 51; DB 14; Length 2485;  
Best Local Similarity 33.3%; Pred. No. 30;  
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 1 MFHISPPKYPHIOEAKKA 21  
:|:::|:| | |:::|  
DB 1212 VFINVRTPKYKHHYQCCEDHA 1232

RESULT 3

ID 09JFT5 PRELIMINARY; PRT; 2485 AA.  
AC 09JFT5;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE NONSTRUCTURAL POLYPROTEIN.  
OS Venezuelan equine encephalitis virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
OC Alphavirus.  
OX NCBI\_TaxID=11036;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=243937;  
RX MEDLINE=99214366; PubMed=10196323;  
RA Wang E., Barrera R., Boshell J., Ferro C., Freter J.E., Navarro J.C.,  
RA Salas R., Vasquez C., Weaver S.C.;  
RT "Genetic and phenotypic changes accompanying the emergence of  
RT enzootic subtype IC Venezuelan equine encephalitis viruses from an  
RT enzootic subtype ID progenitor";  
RL J. Virol. 73:4266-4271(1999).

RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=243937;  
RA Wang E., Weaver S.C.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF004459; AAD43357.1; -.  
DR InterPro; IPR000606; -.  
DR InterPro; IPR002589; -.  
DR InterPro; IPR002620; -.  
DR Pfam; PF01443; Viral\_helicase1; 1.  
DR Pfam; PF01661; DUF27; 1.  
DR Pfam; PF01707; Peptidase\_C9; 1.  
KW Polyprotein.  
SQ SEQUENCE 2485 AA; 277022 MW; AACAT28C25F51C07 CRC64;

Query Match 41.5%; Score 51; DB 14; Length 2485;  
Best Local Similarity 33.3%; Pred. No. 30;  
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 1 MFHISPPKYPHIOEAKKA 21  
:|:::~|:| | |:::~|  
DB 1212 VFINVRTPKYKHHYQCCEDHA 1232

RESULT 4

ID 09IR88 PRELIMINARY; PRT; 2485 AA.  
AC 09IR88;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE NONSTRUCTURAL POLYPROTEIN.  
OS Venezuelan equine encephalitis virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
OC Alphavirus.  
OX NCBI\_TaxID=11036;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SH3;  
RX MEDLINE=99214366; PubMed=10196323;  
RA Wang E., Barrera R., Boshell J., Ferro C., Freter J.E., Navarro J.C.,  
RA Salas R., Vasquez C., Weaver S.C.;  
RT "Genetic and phenotypic changes accompanying the emergence of  
RT enzootic subtype IC Venezuelan equine encephalitis viruses from an  
RT enzootic subtype ID progenitor";  
RL J. Virol. 73:4266-4271(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SH3;  
RA Wang E., Weaver S.C.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U55360; AAD43358.1; -.  
DR InterPro; IPR000606; -.  
DR InterPro; IPR002589; -.  
DR InterPro; IPR002620; -.  
DR Pfam; PF01443; Viral\_helicase1; 1.  
DR Pfam; PF01661; DUF27; 1.  
DR Pfam; PF01707; Peptidase\_C9; 1.  
KW Polyprotein.  
SQ SEQUENCE 2485 AA; 277034 MW; E07E0A68D91140EB CRC64;

Query Match 41.5%; Score 51; DB 14; Length 2485;  
Best Local Similarity 33.3%; Pred. No. 30;  
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 1 MFHISPPKYPHIOEAKKA 21  
:|:::~|:| | |:::~|  
DB 1212 VFINVRTPKYKHHYQCCEDHA 1232

RESULT 5

091CG5 PRELIMINARY: PRT: 2485 AA.  
ID 091CG5  
AC 091CG5  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
DE 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
DE NONSTRUCTURAL POLYPEPTIDE.  
OS Venezuelan equine encephalitis virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
OC Alphavirus.  
OX NCBI\_TaxID=11036;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=66457, AND 66637;  
RX MEDLINE=99214366; PubMed=10196323;  
RA Wang E., Barrera R., Boshell J., Ferro C., Freier J.E., Navarro J.C.,  
RA Salas R., Vasquez C., Weaver S.C.;  
RT "Genetic and phenotypic changes accompanying the emergence of  
RT epizootic subtype IC Venezuelan equine encephalitis viruses from an  
RT epizootic subtype ID progenitor."  
RL J. Virol. 73:4266-4271(1999).  
[2]  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=66457, AND 66637;  
RA Wang E., Weaver S.C.;  
RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF004472; AAD43356.1; -  
DR EMBL; AF004458; AAD43355.1; -  
DR InterPro; IPR000606; -  
DR InterPro; IPR002589; -  
DR InterPro; IPR002620; -  
DR Pfam; PF01661; DUF27.1.  
DR Pfam; PF01707; Peptidase\_C9; 1.  
DR Polyprotein.  
KW POLYPEPTIDE.  
SQ SEQUENCE 2485 AA; 276969 MW; 3F5776AAE8D91F95 CRC64;  
Query Match 41.5%; Score 51; DB 14; Length 2485;  
Best Local Similarity 33.3%; Pred. No. 30;  
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;  
OY 1 MFLHSSPKYPTPHOEAOKA 21  
DB 1212 VF1NVRTPKYKHYYQCEDHA 1232  
RESULT 6  
ID 091R87 PRELIMINARY: PRT: 2492 AA.  
AC 091R87  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
DE 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
DE NONSTRUCTURAL POLYPEPTIDE.  
OS Venezuelan equine encephalitis virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
OC Alphavirus.  
OX NCBI\_TaxID=11036;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=830434;  
RX MEDLINE=99214366; PubMed=10196323;  
RA Wang E., Barrera R., Boshell J., Ferro C., Freier J.E., Navarro J.C.,  
RA Salas R., Vasquez C., Weaver S.C.;  
RT "Genetic and phenotypic changes accompanying the emergence of  
RT epizootic subtype IC Venezuelan equine encephalitis viruses from an  
RT epizootic subtype ID progenitor."  
RL J. Virol. 73:4266-4271(1999).  
[2]  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=830434;  
RA Wang E., Weaver S.C.;

RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL; U55362; AAD43359.1; -  
DR InterPro; IPR000606; -  
DR InterPro; IPR002589; -  
DR InterPro; IPR002620; -  
DR Pfam; PF01443; Viral\_helicase1; 1.  
DR Pfam; PF01661; DUF27.1.  
DR Pfam; PF01707; Peptidase\_C9; 1.  
DR Polyprotein.  
KW POLYPEPTIDE.  
SQ SEQUENCE 2492 AA; 277904 MW; E5DE32A5D2D7A3F3 CRC64;  
Query Match 41.5%; Score 51; DB 14; Length 2492;  
Best Local Similarity 33.3%; Pred. No. 30;  
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;  
OY 1 MFLHSSPKYPTPHOEAOKA 21  
DB 1212 VF1NVRTPKYKHYYQCEDHA 1232  
RESULT 7  
ID 092RY3 PRELIMINARY: PRT: 740 AA.  
AC 092RY3  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)  
DE 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
DE NDX2 HOMEBOX PROTEIN (FRAGMENT).  
GN NDX2.  
OS Lotus japonicus.  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
OC Fabales; Fabaceae; Papilionoideae; Lotus.  
OX NCBI\_TaxID=34305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Jorgensen J.E., Gronlund M., Palisgaard N., Larsen K., Marcker K.A.,  
RA Jensen E.;  
RT "A new class of plant homeobox genes is expressed in specific regions  
RT of determinate symbiotic root nodules."  
RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AJ011829; CAA09792.1; -  
DR InterPro; IPR001356; -  
DR PROSITE; PSS0071; HOMEBOX\_2; 1.  
DR SMART; SM00389; HOX; 1.  
KW Homeobox; DNA-binding; Nuclear protein.  
FT NON\_TER  
SQ SEQUENCE 740 AA; 83081 MW; A3D719B8DCFAFC11 CRC64;  
Query Match 40.7%; Score 50; DB 10; Length 740;  
Best Local Similarity 64.7%; Pred. No. 13;  
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
OY 3 LHSSPKYPTPHOEAOK 19  
DB 438 LFSAGFKONHVOEAOK 454  
RESULT 8  
ID 09WUJ1 PRELIMINARY: PRT: 1879 AA.  
AC 09WUJ1  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)  
DE 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
DE PUTATIVE NONSTRUCTURAL POLYPEPTIDE PRECURSOR.  
GN NS.  
OS Venezuelan equine encephalitis virus (strain Trinidad donkey).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
OC Alphavirus.  
OX NCBI\_TaxID=11038;

KP SEQUENCE FROM N.A.  
 RC STRAIN=MUCCAMBO BEAN 8;  
 RC MEDLINE=99101297; PubMed=9886206;  
 RA Kinney R.M., Pfeiffer M., Tsuchiya K.R., Chang G.J., Roenigk J.T.;  
 RT "Nucleotide sequences of the 26S mRNAs of the viruses defining the  
 RT Venezuelan equine encephalitis antigenic complex.";  
 RN Am. J. Trop. Med. Hyg. 59:952-964(1998).  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MUCCAMBO BEAN 8;  
 RA Kinney R.M., Pfeiffer M., Weissner J.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBD databases.  
 DR EMBL; AF075253; AAD14554.1; -  
 DR InterPro: IPR000606; -  
 DR InterPro: IPR002589; -  
 DR InterPro: IPR002620; -  
 DR InterPro: IPR01443; Viral\_helicase1; 1.  
 DR Pfam: PF01661; DUF27.1.  
 DR Pfam: PF01707; Peptidase\_C9; 1.  
 KW Polypeptin.  
 KW CHAIN 1 535 METHYLTRANSFERASE NSP1.  
 FT CHAIN 536 1329 NONSTRUCTURAL PROTEINASE NSP2.  
 FT CHAIN 1330 1848 NONSTRUCTURAL PROTEIN NSP3.  
 FT CHAIN 1849 2455 RNA POLYMERASE NSP4.  
 SQ SEQUENCE 2455 AA; 273902 MW; 0DDFCB42A2809E52 CRC64;

```
QY      1 MFLHISSPFKYPHTQEAQKEA 21
          :::::|:|:|:|:|:|
Db      1212 VEVNVRTPKYHHYQQCEDHA 1232
```

Query Match	40.7%	Score 50;	DB 14;	Length 1879;
Best Local Similarity	33.3%	Pred. No. 33;		
Matches	7;	Conservative	8;	Mismatches 6; Indels 0; Gaps 0;
QY	1	MFLHISSEPKYRPHTEAQKEA	21	
Db	1212	IFVNVRTPKYKHHNYOCCEDHA	1232	

```

Oy      1  MELHISPEKYPHTOEAQKEA  21
        : : : : : | : : |
Db      1212  VFVANNRTPKYKHNYQCEDEHA  1232

RESULT  10
066594
ID      066594      PRELIMINARY;      PRT:  2493  AA.
AC      066594:
DT      01-NOV-1996  (TREMBLrel. 01. Created)
DT      01-NOV-1996  (TREMBLrel. 01. Last sequence update)
DT      01-MAR-2001  (TREMBLrel. 16. Last annotation update)
DE      NONSTRUCTURAL POLYPEPTIDE.
OS      Venezuelan equine encephalitis virus (strain TC-83).
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OX      Alphavirus.
RN      NCBL_TaxID=11037;
RP      [1]
RR      SEQUENCE FROM N.A.
RC      STRAIN=TC-83;
RX      MEDLINE=86306669; PubMed=3755750;
RA      Johnson B.J., Kinney R.M., Kost C.L., Trent D.W.;
RT      "Molecular determinants of alphavirus neurovirulence: nucleotide and
RT      deduced protein sequence changes during attenuation of Venezuelan
RT      equine encephalitis virus."
RL      J Gen. Virol. 67:1951-1960(1986).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=TC-83;
RX      MEDLINE=89243175; PubMed=2524126;
RA      Kinney R.M., Johnson B.J., Welch J.B., Tsuchiya K.R., Trent D.W.;
RT      "The full-length nucleotide sequences of the virulent Trinidad donkey
RT      strain of Venezuelan equine encephalitis virus and its attenuated
RT      vaccine derivative, strain TC-83."
RL      Virology 170:19-30(1989).
DR      EMBL; L01443; AAB02516.1; -.
DR      InterPro; IPR000606; -.
DR      InterPro; IPR002589; -.
DR      pfam; PF01443; Viral_helicase1; 1.
DR      pfam; PF01661; DUF27; 1.
DR      pfam; PF01707; peptidase_C9; 1.

```

KW Polypeptide. 1 535 POTENTIAL.  
FT CHAIN 536 1329 POTENTIAL.  
FT CHAIN 1330 1886 POTENTIAL.  
FT CHAIN 1887 2493 POTENTIAL.  
SQ SEQUENCE 2493 AA; 277927 MW; 65388F048664796D CRC64;

Query Match 40.7%; Score 50; DB 14; Length 2493;  
Best Local Similarity 33.3%; Pred. No. 43;  
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 1 MFLHISPPKYPHTOEAOKEA 21  
Db 1212 IFVNWRTPKYHHYOCEDHA 1232

RESULT 11  
090163 PRELIMINARY; PRT: 2493 AA.  
AC 090163;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE NONSTRUCTURAL POLYPEPTIDE.  
OS Venezuelan equine encephalitis virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
OC Alphavirus.  
OX NCBI\_TaxID=11036;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-71-180, 600035-71-180/4;  
RX MEDLINE=89243175; PubMed=2524126;  
RA Kinney R.M., Johnson B.J., Welch J.B., Tsuchiya K.R., Trent D.W.;  
RT "The full-length nucleotide sequences of the virulent Trinidad donkey  
RT strain of Venezuelan equine encephalitis virus and its attenuated  
RT vaccine derivative, strain TC-83."  
RL Virology 170:19-30(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-71-180, 600035-71-180/4;  
RX MEDLINE=93107872; PubMed=1469368;  
RA Kinney R.M., Tsuchiya K.R., Sneider J.M., Trent D.W.;  
RT "Molecular evidence for the origin of the widespread Venezuelan equine  
RT encephalitis epidemic of 1969 to 1972."  
RL J. Gen. Virol. 73:0-0(0).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-71-180, 600035-71-180/4;  
RX kinney R.M., Trent D.W.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF069903; AAC24033.1;  
DR InterPro: IPR000606;  
DR InterPro: IPR002589;  
DR InterPro: IPR002620;  
DR Pfam: PF01443; Viral\_helicase1; 1.  
DR Pfam: PF01661; DUF27; 1.  
DR Pfam: PF01707; Peptidase\_C9; 1.  
KW Polypeptide.  
FT CHAIN 1 535 NONSTRUCTURAL PROTEIN NSP1.  
FT CHAIN 536 1329 NONSTRUCTURAL PROTEIN NSP2.  
FT CHAIN 1330 1886 NONSTRUCTURAL PROTEIN NSP3.  
FT CHAIN 1887 2493 NONSTRUCTURAL PROTEIN NSP4.  
SQ SEQUENCE 2493 AA; 277876 MW; 9791EE7D2694A466 CRC64;

Query Match 40.7%; Score 50; DB 14; Length 2493;  
Best Local Similarity 33.3%; Pred. No. 43;  
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 1 MFLHISPPKYPHTOEAOKEA 21  
Db 1212 IFVNWRTPKYHHYOCEDHA 1232

RESULT 12  
066592 PRELIMINARY; PRT: 2493 AA.  
ID 066592;  
AC 066592;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE PUTATIVE NONSTRUCTURAL POLYPEPTIDE PRECURSOR.  
OS Venezuelan equine encephalitis virus (strain Trinidad donkey).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
OC Alphavirus.  
OX NCBI\_TaxID=11038;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TRINIDAD DONKEY;  
RX MEDLINE=86263392; PubMed=3088830;  
RA Kinney R.M., Johnson B.J., Brown V.L., Trent D.W.;  
RT "Nucleotide sequence of the 26 S mRNA of the virulent Trinidad donkey  
RT strain of Venezuelan equine encephalitis virus and deduced sequence of  
RT the encoded structural proteins."  
RL Virology 152:400-413(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TRINIDAD DONKEY;  
RX MEDLINE=86306669; PubMed=3755750;  
RA Johnson B.J., Kinney R.M., Kost C.L., Trent D.W.;  
RT "Molecular determinants of alphavirus neurovirulence: nucleotide and  
RT deduced protein sequence changes during attenuation of Venezuelan  
RT equine encephalitis virus."  
RL J. Gen. Virol. 67:0-0(0).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TRINIDAD DONKEY;  
RX MEDLINE=89243175; PubMed=2524126;  
RA Kinney R.M., Johnson B.J., Welch J.B., Tsuchiya K.R., Trent D.W.;  
RT "The full-length nucleotide sequences of the virulent Trinidad donkey  
RT strain of Venezuelan equine encephalitis virus and its attenuated  
RT vaccine derivative, strain TC-83."  
RL Virology 170:19-30(1989).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TRINIDAD DONKEY;  
RA Kinney R.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: L01442; AAC19321.2;  
DR InterPro: IPR000606;  
DR InterPro: IPR002589;  
DR InterPro: IPR002620;  
DR Pfam: PF01443; Viral\_helicase1; 1.  
DR Pfam: PF01661; DUF27; 1.  
DR Pfam: PF01707; Peptidase\_C9; 1.  
KW Polypeptide.  
FT CHAIN 1 535 NONSTRUCTURAL PROTEIN NSP1.  
FT CHAIN 536 1329 NONSTRUCTURAL PROTEIN NSP2.  
FT CHAIN 1330 1886 NONSTRUCTURAL PROTEIN NSP3.  
FT CHAIN 1887 2493 NONSTRUCTURAL PROTEIN NSP4.  
SQ SEQUENCE 2493 AA; 277929 MW; 7832C85B6D6C0BC CRC64;

Query Match 40.7%; Score 50; DB 14; Length 2493;  
Best Local Similarity 33.3%; Pred. No. 43;  
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 1 MFLHISPPKYPHTOEAOKEA 21  
Db 1212 IFVNWRTPKYHHYOCEDHA 1232

RESULT 13  
O9WJG9

ID Q9WJC9 PRELIMINARY; PRT: 2501 AA.  
AC Q9WJC9;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE NONSTRUCTURAL POLYPEPTIDE.  
OS Venezuelan equine encephalitis virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
OC Alphavirus.  
NCBI\_TaxID=11036;  
OX [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-TONATE CAAN 410D;  
RC MEDLINE=99101297; PubMed=9886206;  
RA Kliney R.M., Pfeiffer M., Tsuchiya K.R., Chang G.J., Roehrig J.T.;  
RT "Nucleotide sequences of the 26S mRNAs of the viruses defining the  
RL Venezuelan equine encephalitis antigenic complex.";  
RL Am. J. Trop. Med. Hyg. 59:952-964(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TONATE CAAN 410D;  
RA Kliney R.M., Pfeiffer M., Meisner J.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF015254; AAD14556.1; -  
DR InterPro; IPR000606; -  
DR InterPro; IPR002589; -  
DR InterPro; IPR002620; -  
DR Pfam; PF01443; Viral\_helicase1; 1.  
DR Pfam; PF01661; DUF27; 1.  
DR Pfam; PF01707; Peptidase\_C9; 1.  
DR Polyprotein.  
FT CHAIN 1 535 METHYLTRANSFERASE NSP1.  
FT CHAIN 536 1329 NONSTRUCTURAL PROTEINASE NSP2.  
FT CHAIN 1330 1894 NONSTRUCTURAL PROTEIN NSP3.  
FT CHAIN 1895 2501 RNA POLYMERASE NSP4.  
SQ SEQUENCE 2501 AA; 278519 MW; 8EAC48DA8CF0E28C CRC64;

Query Match 40.7%; Score 50; DB 14; Length 2501;  
Best Local Similarity 33.3%; Pred. No. 44;  
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 1 MFLHSPPKYPHTQEAQEA 21  
Db 1212 VFVNVRTPEKYHHYQCCEDHA 1232

RESULT 14  
ID Q9NBE9 PRELIMINARY; PRT: 311 AA.  
AC Q9NBE9;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE HYPOTHETICAL 34.7 KDA PROTEIN.  
GN CH1.365.  
OS Trypanosoma brucei.  
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OC NCBI\_TaxID=5691;  
OX [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-TREU927;  
RA Hall N., Bowman S., Quail M., Ivens A.C., Kay M.P., Bray-Allen S.,  
RA Lennard N.J., Clark L.N., Harris B.R., Melville S., Lawson D.,  
RA Gerrard C., Rajandream M.A., Barrell B.G.;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL359782; CAB95594.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 311 AA; 34735 MW; 726BD2A34459276E CRC64;

Query Match 38.2%; Score 47; DB 5; Length 311;  
Best Local Similarity 44.4%; Pred. No. 15;

Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;  
OY 6 SSPEYPTHTQEAQEAQR 23  
Db 48 SHPSHNHNDQDQDQDQR 65

RESULT 15  
ID Q9N801 PRELIMINARY; PRT: 1747 AA.  
AC Q9N801;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE POSSIBLE T16011.22 PROTEIN.  
GN CH1.202.  
OS Trypanosoma brucei.  
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OC NCBI\_TaxID=5691;  
OX [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-TREU927;  
RA Hall N., Bowman S., Quail M., Ivens A.C., Kay M.P., Bray-Allen S.,  
RA Lennard N.J., Clark L.N., Harris B.R., Melville S., Lawson D.,  
RA Gerrard C., Rajandream M.A., Barrell B.G.;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL359782; CAB95482.1; -  
DR InterPro; IPR001680; -  
SQ SEQUENCE 1747 AA; 187999 MW; 6E21EEFEC796824 CRC64;

Query Match 38.2%; Score 47; DB 5; Length 1747;  
Best Local Similarity 72.7%; Pred. No. 89;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 LHISPPKYPH 13  
Db 1612 LHISPPCYCPH 1632

Search completed: October 16, 2001, 19:07:02  
Job time: 300 sec





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 16, 2001, 19:02:37 ; Search time 33.92 Seconds  
(without alignments)  
23.227 Million cell updates/sec

Title: US-09-675-650-3  
Perfect score: 123  
Sequence: 1 MFLHSSPFKYPHQAQKEAQR 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues  
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	41.5	2485	1	POLN.EEVV3
2	51	41.5	2492	1	POLN.EEVV3
3	50	40.7	2492	1	POLN.EEVV3
4	46.5	37.8	719	1	TE80.TETTH
5	45.5	37.0	580	1	5NPD.BOOMI
6	44.5	36.2	511	1	VGIG.VSVIG
7	44.5	36.2	511	1	VGIG.VSVIG
8	44.5	36.2	511	1	VGIG.VSVIG
9	44.5	36.2	511	1	VGIG.VSVIG
10	44	35.8	247	1	YEPD.YERPE
11	44	35.8	297	1	YEPD.YERPE
12	44	35.8	544	1	YEPD.YERPE
13	44	35.8	774	1	SCA.DROME
14	43.5	35.4	258	1	KLKI.PAPHA
15	43.5	35.4	262	1	KLKI.PAPHA
16	43	35.0	628	1	V7OK.TYVW
17	43	35.0	628	1	V7OK.TYVW
18	43	35.0	628	1	V7OK.TYVW
19	42.5	34.6	526	1	MITF.MOUSE
20	42	34.1	201	1	IORB.ARCFU
21	42	34.1	379	1	NIR.PESP
22	42	34.1	395	1	VIBC.VIBCH
23	41.5	33.7	432	1	CGA2.HUMAN
24	41.5	33.7	499	1	CPML.ONCMT
25	41.5	33.7	1135	1	PHYC.SORBI
26	41.5	33.7	1137	1	PHYC.SORBI
27	41	33.3	412	1	ISPL.HUMAN
28	41	33.3	497	1	HMS.DROME
29	41	33.3	905	1	RPA1.THECE
30	40.5	32.9	257	1	KLKI.MACFA
31	40.5	32.9	491	1	CPB4.RABIT
32	40.5	32.9	491	1	CPB5.RABIT
33	40.5	32.9	1005	1	BGAL.ACTPL

## ALIGNMENTS

RESULT 1	ID	Score	Query Match	Length	DB ID	Description
AC	POLN.EEVV3	51	41.5	2485	1	POLN.EEVV3
AC	POLN.EEVV3	51	41.5	2492	1	POLN.EEVV3
AC	POLN.EEVV3	50	40.7	2492	1	POLN.EEVV3
AC	POLN.EEVV3	46.5	37.8	719	1	TE80.TETTH
AC	POLN.EEVV3	45.5	37.0	580	1	5NPD.BOOMI
AC	POLN.EEVV3	44.5	36.2	511	1	VGIG.VSVIG
AC	POLN.EEVV3	44.5	36.2	511	1	VGIG.VSVIG
AC	POLN.EEVV3	44.5	36.2	511	1	VGIG.VSVIG
AC	POLN.EEVV3	44.5	36.2	511	1	VGIG.VSVIG
AC	POLN.EEVV3	44	35.8	247	1	YEPD.YERPE
AC	POLN.EEVV3	44	35.8	297	1	YEPD.YERPE
AC	POLN.EEVV3	44	35.8	544	1	YEPD.YERPE
AC	POLN.EEVV3	44	35.8	774	1	SCA.DROME
AC	POLN.EEVV3	43.5	35.4	258	1	KLKI.PAPHA
AC	POLN.EEVV3	43.5	35.4	262	1	KLKI.PAPHA
AC	POLN.EEVV3	43	35.0	628	1	V7OK.TYVW
AC	POLN.EEVV3	43	35.0	628	1	V7OK.TYVW
AC	POLN.EEVV3	43	35.0	628	1	V7OK.TYVW
AC	POLN.EEVV3	42.5	34.6	526	1	MITF.MOUSE
AC	POLN.EEVV3	42	34.1	201	1	IORB.ARCFU
AC	POLN.EEVV3	42	34.1	379	1	NIR.PESP
AC	POLN.EEVV3	42	34.1	395	1	VIBC.VIBCH
AC	POLN.EEVV3	41.5	33.7	432	1	CGA2.HUMAN
AC	POLN.EEVV3	41.5	33.7	499	1	CPML.ONCMT
AC	POLN.EEVV3	41.5	33.7	1135	1	PHYC.SORBI
AC	POLN.EEVV3	41.5	33.7	1137	1	PHYC.SORBI
AC	POLN.EEVV3	41	33.3	412	1	ISPL.HUMAN
AC	POLN.EEVV3	41	33.3	497	1	HMS.DROME
AC	POLN.EEVV3	41	33.3	905	1	RPA1.THECE
AC	POLN.EEVV3	40.5	32.9	257	1	KLKI.MACFA
AC	POLN.EEVV3	40.5	32.9	491	1	CPB4.RABIT
AC	POLN.EEVV3	40.5	32.9	491	1	CPB5.RABIT
AC	POLN.EEVV3	40.5	32.9	1005	1	BGAL.ACTPL

OY	1	MEHLISSPEKPYPHTOEAOKEA	21	:	:
DB	1212	IFINWRTPKYKHHYQCCEDHA	1232		
	RESULT	2			
	POLN_EEVPV	STANDARD;	PRT;	2492	AA.
AC	P36328;				
DT	01-JUN-1994 (Rel. 29, Created)				
DT	01-JUN-1994 (Rel. 29, Last sequence update)				
DT	01-OCT-1996 (Rel. 34, Last annotation update)				
DE	NSP4]				
OS	Venezuelan equine encephalitis virus (strain P676).				
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;				
OC	Alphavirus.				
OX	NCBI_TaxID=36385;				
RN	[1]				
RX	SEQUENCE FROM N.A.				
RA	MEDLINE=93079859; PubMed=1448915;				
RT	Kliney R.M., Tsuchiya K.R., Sneller J.M., Trent D.W.:				
RT	*Genetic evidence that epizootic Venezuelan equine encephalitis (VEE)				
RL	viruses may have evolved from enzootic VEE subtype I-D virus.*;				
RL	Virology 191:569-580(1992).				
CC	-I- FUNCTION: NSP2 MAY BE INVOLVED IN RNA BINDING DURING REPLICATION.				
CC	-I- PPM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATRE PROTEINS.				
CC	-I- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UGA OCCURS				
CC	BETWEEN THE CODONS FOR 1879-GLN AND 1880-ARG.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@lsb-sib.ch">license@lsb-sib.ch</a> ).				
CC	-----				
DR	EMBL; L04653; CAB23722.1; -				
DR	PIR; A44213; A44213.				
DR	MEROPS; C09.001; -				
DR	InterPro; IPR000606; -				
DR	InterPro; IPR001788; -				
DR	InterPro; IPR002589; -				
DR	InterPro; IPR002620; -				
DR	Pfam; PF01661; DUF27.1.				
DR	Pfam; PF01707; Peptidase_C9.1.				
DR	Pfam; PF00978; RNA_dep_RNApol2.2.				
DR	Pfam; PF01443; Viral_helicase1.1.				
KW	Polypeptide; Nonstructural protein; RNA-binding; ATP-binding;				
KW	Helicase.				
FT	CHAIN 1 535 NONSTRUCTURAL PROTEIN NSP1.				
FT	CHAIN 536 1329 NONSTRUCTURAL PROTEIN NSP2.				
FT	CHAIN 1330 1679 NONSTRUCTURAL PROTEIN NSP3.				
FT	CHAIN 1680 2492 NONSTRUCTURAL PROTEIN NSP4.				
FT	NP_BIND 721 728 ATP (POTENTIAL).				
FT	SEQUENCE 2492 AA; 277836 MW; E55D056CD7AEDEE CRC64;				
QY	1	MEHLISSPEKPYPHTOEAOKEA	21	:	:
DB	1212	VEINWRTPKYKHHYQCCEDHA	1232		
	Query Match	41.5%;	Score 51;	DB 1;	Length 2492;
	Best Local Similarity	33.3%;	Pred. No. 9.9;		
	Matches 7; Conservative	8;	Mismatches	6;	Indels 0;
					Gaps 0.
	RESULT	3			
	POLN_EEVPV	STANDARD;	PRT;	2492	AA.
ID	POLN_EEVPV				
AC	P27282;				

```

DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE NONSTRUCTURAL POLYPEPTIDE [CONTAINS: NONSTRUCTURAL PROTEINS NSP1 TO NSP4].
OS Venezuelan equine encephalitis virus (strain Trinidad donkey).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae; Alphavirus.
OC NCBI_TaxID=11038;
OX [1]
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=89243175; PubMed=2524126;
RA Kinney R.M., Johnson B.J.B., Welch J.B., Teuchija K.R., Trent D.W.;
RT "The full-length nucleotide sequences of the virulent Trinidad donkey strain of Venezuelan equine encephalitis virus and its attenuated vaccine derivative, strain TC-83."
RL Virology 170:119-30(1989).
CC -1- FUNCTION: NSP2 MAY BE INVOLVED IN RNA BINDING DURING REPLICATION.
CC -1- PWM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UGA OCCURS BETWEEN THE CODONS FOR 1879-GLN AND 1880-ARG.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04332; AAB02518.1; -.
DR PIR: A31467; MNWVTD.
DR MEROPS: C09.001; -.
DR InterPro: IPR000606; -.
DR InterPro: IPR002589; -.
DR InterPro: IPR002620; -.
DR Pfam: PF01661; DUF27.1.
DR Pfam: PF01707; Peptidase_C9.1.
DR Pfam: PF01443; Viral_helicase1.1.
DR Polyprotein; Nonstructural protein; RNA-binding; ATP-binding; Helicase.
KW CHAIN 1 535 NONSTRUCTURAL PROTEIN NSP1.
FT CHAIN 536 1329 NONSTRUCTURAL PROTEIN NSP2.
FT CHAIN 1330 1879 NONSTRUCTURAL PROTEIN NSP3.
FT CHAIN 1880 2492 NONSTRUCTURAL PROTEIN NSP4.
FT NP_BIND 721 728 ATP (POTENTIAL).
SQ SEQUENCE 2492 AA; 277902 MW; 1BAD415B70DC3FA0 CRC64;

Query Match 40.7%; Score 50; DB 1; Length 2492;
Best Local Similarity 33.3%; Pred. No. 14;
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0.

Oy 1 MFLHTSSPKYPTHTOEAKKA 21
Db 1212 IFVNWRTPKYHHYQCEDHA 1232

RESULT 4
TE80_TETTH STANDARD: PRT; 719 AA.
AC Q94818;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE TELOMERASE COMPONENT P80 (EC 2.7.7.-).
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida; Tetrahymenida; Tetrahymena.
OX NCBI_TaxID=5911;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=9592335; PubMed=7774009;

```

```

RA Collins K., Kobayashi R., Gredele C.W.:  

RT "Purification of Tetrahymena telomerase and cloning of genes encoding  

RL the two protein components of the enzyme."; Cell 81:677-686(1995).  

CC -1- FUNCTION: RIBONUCLEOPROTEIN DNA POLYMERASE THAT CATALYZES THE DE  

CC NOVO SYNTHESIS OF TELOMERIC SIMPLE SEQUENCE REPEATS. P80 BINDS  

CC TIGHTLY AND SPECIFICALLY TO THE TELOMERASE RNA SUGGESTING ITS  

CC ASSOCIATION WITH A REGION OF RNA SECONDARY STRUCTURE.  

CC -1- SUBUNIT: TELOMERASE CONSIST OF TWO SUBUNIT, P80 AND P95 THAT FORM  

CC A 1:1:1 COMPLEX WITH THE 159 NT TELOMERASE RNA.  

CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  

CC -----  

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  

CC the European Bioinformatics Institute. There are no restrictions on its  

CC use by non-profit institutions as long as its content is in no way  

CC modified and this statement is not removed. Usage by and for commercial  

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  

CC or send an email to license@isb-sib.ch).  

CC -----  

DR EMBL; U23641; AAC46601.1; -  

RW Translase; RNA-directed DNA polymerase; Telomere; Nuclear protein;  

KW DNA-binding.  

SQ SEQUENCE 719 AA; 82351 MW; 8A945A71189CA99C CRC64;  

  

QY 2 FLHSSP-----FKYPTQEAQKKA 21  

Db 298 FCHISEPKERYKTKIGKKYKPTTEYKKA 326  

  

RESULT 5  

5NTD_BOOMT STANDARD; PRT; 580 AA.  

AC P52307; P90696;  

DT 01-OCT-1996 (Rel. 34, Created)  

DT 01-OCT-2000 (Rel. 40, Last sequence update)  

DT 01-OCT-2000 (Rel. 40, Last annotation update)  

DE 5'-NUCLEOTIDASE PRECURSOR (EC 3.1.3.5) (FRAGMENT).  

OS Boophilus microplus (Cattle tick).  

OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  

OC Parasitiformes; Ixodida; Ixodidae; Boophilus.  

OX NCBI_Taxid=6941;  

RN [1]  

RP SEQUENCE FROM N.A.  

RA MEDLINE=99306777; PubMed=10380109;  

RA Llyou N., Hamilton S., Elyin C., Willadsen P.;  

RT "Cloning and expression of ecto 5-nucleotidase from the cattle tick  

RT Boophilus microplus."; Insect Mol. Biol. 8:257-266(1999).  

RL [2]  

RN SEQUENCE OF 15-40 AND 162-180.  

RX MEDLINE=93250870; PubMed=8387372;  

RA Willadsen P., Riding G.A., Jarney J., Atkins A.;  

RT "The nucleotidase of boophilus microplus and its relationship to  

RT enzymes from the rat and Escherichia coli."; Insect Biochem. Mol. Biol. 23:291-295(1993).  

RL -1- FUNCTION: DEGRADATION OF EXTERNAL UDP-GLUCOSE TO URIDINE  

CC MONOPHOSPHATE AND GLUCOSE-1-PHOSPHATE, WHICH CAN THEN BE USED BY  

CC THE CELL.  

CC -1- CATALYTIC ACTIVITY: UDP-SUGAR + H(2)O =UMP + SUGAR 1-PHOSPHATE.  

CC -1- CATALYTIC ACTIVITY: A 5'-RIBONUCLEOTIDE + H(2)O = A  

CC RIBONUCLEOSIDE + ORTHOPHOSPHATE.  

CC -1- COFACTOR: ZINC.  

CC -1- SUBUNIT: HOMODIMER.  

CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  

CC -1- TISSUE SPECIFICITY: GUT, OVARIES AND SALIVARY GLANDS.  

CC -1- PMW: GLYCOSYLATED.  

CC -1- SIMILARITY: BELONGS TO THE 5'-NUCLEOTIDASE FAMILY.

```

CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-
CC	EMBL; U80634; AAB38963.1; -.
DR	InterPro; IPR000934; -.
DR	InterPro; IPR002224; -.
DR	PROSITE; PS00785; 5_NUCLEOTIDASE_1; FALSE_NEG.
DR	PROSITE; PS00786; 5_NUCLEOTIDASE_2; FALSE_NEG.
DR	Pfam; PF01009; 5_nucleotidase; 1.
KW	Hydrolase; GPI-anchor; Glycoprotein; signal; zinc.
FT	NON_TER 1 1
FT	SIGNAL <1 14
FT	CHAIN 15 552
FT	PROPEP 553 580
FT	LIPID 552 552
FT	CARBOHYD 172 172
FT	CARBOHYD 285 285
FT	CARBOHYD 423 423
FT	CARBOHYD 536 536
FT	CONFLICT 15 15
FT	CONFLICT 37 39
SO	SEQUENCE 580 AA; 63460 MW; 588EEF2014071A87 CRC64;
OY	4 HISSEFKYPHTQEA---OKEAOR 23
DB	175 YLSPGKVFRTDEACIQREAOR 197
VGIG_VSYIG	RESULT 6 VGIG_VSYIG STANDARD; PRT; 511 AA.
AC	P04883;
DT	13-AUG-1987 (Rel. 05, Created)
DT	13-AUG-1987 (Rel. 05, Last sequence update)
DT	01-OCT-1994 (Rel. 30, Last annotation update)
DE	SPIKE GLYCOPROTEIN PRECURSOR.
GN	G.
OS	Vesicular stomatitis virus (serotype Indiana / strain Glasgow).
OC	Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC	Rhabdoviridae; Vesiculovirus.
OX	NCBI_TaxId=11278;
RN	[1]
RF	SEQUENCE FROM N.A.
RX	MEDLINE=86142630; PubMed=3005478;
RA	Vandepool S.B., Holland J.J.;
RT	"Evolution of vesicular stomatitis virus in athymic nude mice: mutations associated with natural killer cell selection.";
RL	J. Gen. Virol. 67:441-451(1986).
CC	-1- FUNCTION: THIS PROTEIN FORMS SPIKES ON THE SURFACE OF THE VIRION. IT IS RESPONSIBLE BOTH FOR THE BINDING OF THE VIRUS TO SUSCEPTIBLE HOST CELLS AND FOR INDUCING THE UPTAKE OF THE VIRUS BY THE CELL. THE INTERACTION BETWEEN THE INTERNAL COMPONENTS OF THE VIRION AND THE PORTION OF THE GLYCOPROTEIN EXPOSED ON THE CYTOPLASMIC FACE OF THE PLASMA MEMBRANE PROBABLY DIRECTS ENVELOPMENT AND VIRUS BUDDING.
CC	-1- PTM: MODIFIED BY THE COVALENT ADDITION OF PALMITIC ACID VIA A THIOETHER LINKAGE TO A CYSTEINE.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).



OY 3 LHSSP---FKYPTPOEAKKE 20  
 DB 412 LHSSKAOYFEPHIDODASQ 432

RESULT 9  
 MTFE\_HUMAN STANDARD: PRT: 526 AA.  
 ID MTFE\_HUMAN 075030; Q14841; Q9P2V0; Q9P2V2; Q9P2Y8;  
 AC 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE MICROPHthalmia-ASSOCIATED TRANSCRIPTION FACTOR.  
 GN MTFE.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS A AND H).  
 RC TISSUE=Kidney;  
 RX MEDLINE=98321192; PubMed=9647758;  
 RA Amae S., Fuse N., Yasumoto K.-I., Sato S., Tajima I., Yamamoto H.,  
 RA Udono T., Duriu Y.K., Tamai M., Takahashi K., Shibahara S.;  
 RT "Identification of a novel isoform of microphthalmia-associated  
 RT transcription factor that is enriched in retinal pigment  
 RT epithelium.";  
 RL Biochem. Biophys. Res. Commun. 247:710-715(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM M).  
 RC TISSUE=Skin;  
 RX MEDLINE=94348499; PubMed=8069297;  
 RA Tachibana M., Perez-Jurado L.A., Nakayama A., Hodgkinson C.A., Li X.,  
 RA Schneider M., Miki T., Fex J., Franke U., Arnhelster H.;  
 RT "Cloning of MTF, the human homolog of the mouse microphthalmia gene  
 RT and assignment to chromosome 3p14.1-p12.3.";  
 RL Hum. Mol. Genet. 3:553-557(1994).  
 RN [3]  
 RP SEQUENCE OF 1-11 FROM N.A. (ISOFORM M).  
 RC TISSUE=Skin;  
 RX MEDLINE=98160190; PubMed=9500554;  
 RA Watanabe A., Takeda K., Ploplis B., Tachibana M.;  
 RT "Epistatic relationship between Waardenburg syndrome genes MTF and  
 RT PAX3.";  
 RL Nat. Genet. 18:283-286(1998).  
 RN [4]  
 RP SEQUENCE OF 1-130 FROM N.A. (ISOFORM C).  
 RC TISSUE=Kidney;  
 RX MEDLINE=20047057; PubMed=10578055;  
 RA Fuse N., Yasumoto K.-I., Takeda K., Amae S., Yoshitawa M., Udono T.,  
 RA Takahashi K., Tamai M., Tomita Y., Tachibana M., Shibahara S.;  
 RT "Molecular cloning of cDNA encoding a novel microphthalmia-associated  
 RT transcription factor isoform with a distinct amino-terminus.";  
 RL J. Biochem. 126:1043-1051(1999).  
 RN [5]  
 RP PARIAL SEQUENCE FROM N.A. (ISOFORMS A; B; H AND M).  
 RX MEDLINE=20223461; PubMed=10760582;  
 RA Udono T., Yasumoto K.-I., Takeda K., Amae S., Watanabe K.-I.,  
 RA Saito H., Fuse N., Tachibana M., Takahashi K., Tamai M., Shibahara S.;  
 RT "Structural organization of the human microphthalmia-associated  
 RT transcription factor gene containing four alternative promoters.";  
 RL Biochim. Biophys. Acta 1491:205-219(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM A).  
 RC TISSUE=Uterus;  
 RX Koehler K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;  
 RT Submitted (Aug-1999) to the EMBL/GenBank/DBD databases.  
 RN [7]  
 RP MUTAGENESIS, PHOSPHORYLATION SITE SER-405, AND DOMAINS.  
 RX MEDLINE=20056131; PubMed=10587587;  
 RA Takeda K., Takemoto C., Kobayashi I., Watanabe A., Nobukuni Y.,  
 RA Fisher D.E., Tachibana M.;

RT Ser298 of MTF, a mutation site in Waardenburg syndrome type 2, is a  
 RT phosphorylation site with functional significance.";  
 RT Hum. Mol. Genet. 9:125-132(2000).  
 RN [8]  
 RP VARIANTS WS2 LYS-310; ARG-324 DEL; PRO-357; ASP-385 AND PRO-405.  
 RX MEDLINE=96154685; PubMed=8589691;  
 RA Tassabehji M., Newton J.E., Liu X.-Z., Brady A., Donnal D.,  
 RA Krajewska-Walasek M., Murday V., Norman A., Obezylyn E., Reardon W.,  
 RA Rice J.C., Trembath R., Wieacker P., Whiteford M., Winter R.,  
 RA Read A.P.;  
 RT "The mutational spectrum in Waardenburg syndrome.";  
 RL Hum. Mol. Genet. 4:2131-2137(1995).  
 RN [9]  
 RP VARIANT TIETZ SYNDROME LYS-317.  
 RC TISSUE=Blood;  
 RX MEDLINE=20311381; PubMed=10851256;  
 RA Smith S.D., Kelley P.M., Kenyon J.B., Hoover D.;  
 RT "Tietz syndrome (hypopigmentation/deafness) caused by mutation of  
 RT MTF.";  
 RL J. Med. Genet. 37:446-448(2000).  
 CC -1- FUNCTION: TRANSCRIPTION FACTOR FOR TYROSINASE AND TYROSINASE-  
 CC RELATED PROTEIN 1. BINDS TO A SYMMETRICAL DNA SEQUENCE (E-BOXES)  
 CC (5'-CAGGTG-3') FOUND IN THE TYROSINASE PROMOTER. PLAYS A CRITICAL  
 CC ROLE IN THE DIFFERENTIATION OF VARIOUS CELL TYPES AS NEURAL CREST-  
 CC DERIVED MELANOCYTES, MAST CELLS, OSTEOCLASTS AND OPTIC CUP-DERIVED  
 CC RETINAL PIGMENT EPITHELIUM.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. BINDS DNA IN THE FORM OF HOMODIMER OR HETERODIMER  
 CC WITH EITHER TFEB, TFEB OR TFEC.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- ALTERNATIVE PRODUCTS: 10 ISOFORMS: A1 (SHOWN HERE), A2 B1, B2, C1,  
 CC C2, H1, H2, M1 AND M2; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC THE X2-TYPE ISOFORMS DIFFER FROM THE X1-TYPE BY THE ABSENCE OF  
 CC A 6-RESIDUE INSERT.  
 CC -1- TISSUE SPECIFICITY: ISOFORM M IS EXCLUSIVELY EXPRESSED IN  
 CC MELANOCYTES AND MELANOMA CELLS. ISOFORMS A AND H ARE WIDELY  
 CC EXPRESSED IN MANY CELL TYPES INCLUDING MELANOCYTES AND RETINAL  
 CC PIGMENT EPITHELIUM (RPE). ISOFORM C IS EXPRESSED IN MANY CELL  
 CC TYPES INCLUDING RPE BUT NOT IN MELANOCYTE-LINEAGE CELLS.  
 CC -1- P.TM: PHOSPHORYLATION AT SER-405 SIGNIFICANTLY ENHANCES THE ABILITY  
 CC TO BIND THE TYROSINASE PROMOTER.  
 CC -1- DISEASE: DEFECTS IN MTF ARE THE CAUSE OF WAARDENBURG SYNDROME  
 CC TYPE 2A (WS2 OR WS2A). A DOMINANT INHERITED DISORDER CHARACTERIZED  
 CC BY SENSORINEURAL HEARING LOSS AND PATCHES OF DEPIGMENTATION. THE  
 CC FEATURES SHOW VARIABLE EXPRESSION AND PENETRANCE.  
 CC -1- DISEASE: DEFECTS IN MTF ARE THE CAUSE OF TIETZ SYNDROME, AN  
 CC AUTOSOMAL DOMINANT DISORDER CHARACTERIZED BY GENERALIZED  
 CC HYPOPIGMENTATION AND PROFOUND, CONGENITAL, BILATERAL DEAFNESS.  
 CC PENETRANCE IS COMPLETE.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB006909; BAA32288.1; -;  
 DR EMBL: AB006989; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL: 229678; CAA62775.1; -;  
 DR EMBL: AF034755; AAC39639.1; -;  
 DR EMBL: AB032359; BAA95208.1; -;  
 DR EMBL: AB032358; BAA95207.1; -;  
 DR EMBL: AB032357; BAA95206.1; -;  
 DR EMBL: AB009608; BAA95209.1; -;  
 DR EMBL: AL10195; CAB53672.1; -;  
 DR EMBL: 156845; -;  
 DR MIM: 103470; -;  
 DR MIM: 103500; -;  
 DR MIM: 193510; -;

DR HSSP; P22415; 1AN4.  
 DR InterPro: IPRO01092; -  
 DR InterPro: IPRO03015; -  
 DR Pfam: PF00010; HLH; 1.  
 DR PROSITE: PS00036; HELIX\_LOOP\_HELIX; 1.  
 DR Developmental protein; Transcription regulation; Activator;  
 KW Nuclear protein; DNA-binding; Alternative splicing; Phosphorylation;  
 KW Disease mutation; Deafness.  
 FT DNA\_BIND 224 295  
 FT DNAS\_BIND 309 324  
 FT DOMAIN 325 365  
 FT DOMAIN 374 395  
 FT DOMAIN 401 431  
 FT MOD\_RES 405 405  
 FT VARSPPLIC 1 34  
 FT VARSPPLIC 1 34  
 FT VARSPPLIC 1 34  
 FT VARSPPLIC 1 35  
 FT VARSPPLIC 1 118  
 FT VARSPPLIC 294 299  
 FT VARSPPLIC 310 310  
 FT VARSPPLIC 317 317  
 FT VARSPPLIC 324 324  
 FT VARSPPLIC 357 357  
 FT VARSPPLIC 385 385  
 FT VARSPPLIC 405 405  
 FT MUTAGEN 405 405  
 FT SEQUENCE 526 AA; 58795 MW; 136EBED304C1986 CRC64;  
 Query Match 36.2%; Score 44.5; DB 1; Length 526;  
 Best Local Similarity 47.1%; Pred. No. 21;  
 Matches 8; Conservative 5; Mismatches 3; Indels 1; Gaps 1;  
 Oy 4 HISSPKYPTQEAQKE 20  
 1:| | | | | | | | | |  
 Db 122 HLENPTKY-HIDQAOQRO 137  
 RESULT 10  
 ID C21U\_HUMAN STANDARD; PRT; 247 AA.  
 AC O9UPM2; 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PUTATIVE PROTEIN C21ORF30 (FRAGMENT).  
 GN C21ORF30.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Oltensweiler B., Obermaier B., Mewes H.-W., Gassenhuber J.,  
 Wiemann S.;

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-20289799; PubMed-10830953;  
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,  
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,  
 RA Ohki M., Takagi T., Sakaki Y., Taudien S., Bleeschmidt K., Polley A.,  
 RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,  
 RA Reischwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,  
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,  
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,  
 RA Minoshima S., Shimizu N., Nordstam G., Hornischer K., Brandt P.,  
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,  
 RA Ramser J., Beck A., Klages S., Hennig S., Klesselmann L., Dagand E.,  
 RA Wehrmeyer S., Borzym K., Gardiner K., Mizellic D., Francis F.,  
 RA Lehrach H., Reinhardt R., Yaspo M.-L.,  
 RA "The DNA sequence of human chromosome 21."  
 FT Nature 405:311-319(2000).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL11578; CAB56001.1; -  
 DR EMBL; AP001754; -; NOT\_ANNOTATED\_CDS.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 FT SEQUENCE 247 AA; 26427 MW; EE84E0A0E4BEF4A CRC64;  
 Query Match 35.8%; Score 44; DB 1; Length 247;  
 Best Local Similarity 43.8%; Pred. No. 12;  
 Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
 Oy 2 FLHISPKYPTQEA 17  
 1:| | | | | | | | | |  
 Db 169 FPHLGRPMAGPHTSOA 184  
 RESULT 11  
 ID YFED\_YERPE STANDARD; PRT; 297 AA.  
 AC O56955; 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE CHELATED IRON TRANSPORT SYSTEM MEMBRANE PROTEIN YFED.  
 GN YFED.  
 OS Yersinia pestis.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Yersinia.  
 OX NCBI\_TaxID=632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-KIM6;  
 RA Bearden S.W., Stags T.M., Perry R.D.;  
 RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PART OF AN ATP-DRIVEN TRANSPORT SYSTEM FOR CHELATED  
 CC IRON.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
 CC (PROBABLE).  
 CC -----  
 CC -1- SIMILARITY: BELONGS TO THE ABC-3 SUBFAMILY OF INTEGRAL MEMBRANE  
 CC PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial



```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, U50597; AAC6150.1; -.
DR InterPro: IPR001626; -.
DR Pfam: PF00950; ABC_3; 1.
KW Transport; Transmembrane; Inner membrane.
FT TRANSMEM 20 40 POTENTIAL.
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 96 116 POTENTIAL.
FT TRANSMEM 133 153 POTENTIAL.
FT TRANSMEM 172 192 POTENTIAL.
FT TRANSMEM 197 217 POTENTIAL.
FT TRANSMEM 224 244 POTENTIAL.
FT TRANSMEM 248 268 POTENTIAL.
SQ SEQUENCE 297 AA; 32187 MW; 2D434515C15D46BE CRC64;

Query Match 35.8%; Score 44; DB 1; Length 297;
Best Local Similarity 52.9%; Pred. No. 14;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 MFLHSPEFKYPHTQEA 17
   : 1 1 1 1 1 1
Db 4 LFLSISEFFAYPFMORA 20

RESULT 12
YL33_CAEBL YL33_CAEBL STANDARD; PRT; 544 AA.
AC P34424;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE HYPOTHETICAL 61.8 KDA PROTEIN P44B9.3 IN CHROMOSOME III.
GN P44B9.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STPAIN-BRISTOL N2;
RA MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copesey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favelle A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hiller L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laistner N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Stoden R.,
RA Stulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman R.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on this
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, L23648; AAA28033.1; -.
DR PIR: S44814; S44814.
DR Wormpep: F44B9.3; CE00520.
KW Hypothetical protein.
SQ SEQUENCE 544 AA; 61809 MW; 0DE93A8B07A02039 CRC64;

```

QY	9	FKYPHTOAKE	20	35.8%;	Score 44;	DB 1;	Length 544;			
Db	115	FKYPHNOFOSE	126	Best Local Similarity 58.3%;	Pred. No. 25;	Matches 7;	Conservative 2;	Mismatches 3;	Indels 0;	Gaps 0;
RESULT	13									
SCA_DROME	STANDARD;	PRT;	774 AA.							
AC	P21520;									
DT	01-MAY-1991 (Rel. 18, Created)									
DT	01-MAY-1991 (Rel. 18, Last sequence update)									
DT	01-OCT-2000 (Rel. 40, Last annotation update)									
DE	SCABROUS PROTEIN PRECURSOR.									
GN	SCA.									
OS	Drosophila melanogaster (Fruit fly).									
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;									
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;									
OC	Ephydroidea; Drosophilidae; Drosophila.									
OX	NCBI_TaxID=7227;									
RN	[1]									
RP	SEQUENCE FROM N.A.									
RX	MEDLINE=91075223; PubMed=2175046;									
RA	Baker N.E., Mlodzik M., Rudin G.M.;									
RT	"Spacing differentiation in the developing Drosophila eye: a									
RL	fibrinogen-related lateral inhibitor encoded by scabrous.";									
SC	Science 250:1370-1377(1990).									
CC	-1- FUNCTION: INVOLVED IN REGULATION OF NEUROGENESIS IN DROSOPHILA.									
CC	SCABROUS MAY ENCODE A LATERAL INHIBITOR OF R8 DIFFERENTIATION.									
CC	-1- MISCELLANEOUS: POSSESSES FIVE PAIRS OF DIABASIC RESIDUES.									
CC	-1- SIMILARITY: CONTAINS 1 FIBRINOGEN C-TERMINAL DOMAIN.									
CC	-----									
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration									
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -									
CC	the European Bioinformatics Institute. There are no restrictions on its									
CC	use by non-profit institutions as long as its content is in no way									
CC	modified and this statement is not removed. Usage by and for commercial									
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>									
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).									
CC	-----									
DR	EMBL; M60065; AAA28880.1; -									
DR	PIR; A39832; A39832.									
DR	HSSP; P02671; 1FED.									
DR	FLYBase; FBgn003326; sca.									
DR	InterPro; IPR002181; -									
DR	Pfam; PF00147; fibrinogen_C; 1.									
DR	PROSITE; PS00514; FIBRIN_AG_C_DOMAIN; 1.									
KW	developmental protein; Neurogenesis; Signal.									
FT	SIGNAL	1	22	POTENTIAL.						
FT	CHAIN	23	774	SCABROUS PROTEIN.						
FT	DOMAIN	504	704	FIBRINOGEN C-TERMINAL.						
FT	DISULFID	517	543	BY SIMILARITY.						
FT	DISULFID	662	675	BY SIMILARITY.						
FT	CARBOHYD	347	347	N-LINKED (GLCNAC. . .) (POTENTIAL).						
FT	CARBOHYD	562	562	N-LINKED (GLCNAC. . .) (POTENTIAL).						
FT	CARBOHYD	593	593	N-LINKED (GLCNAC. . .) (POTENTIAL).						
FT	CARBOHYD	635	635	N-LINKED (GLCNAC. . .) (POTENTIAL).						
FT	CARBOHYD	719	719	N-LINKED (GLCNAC. . .) (POTENTIAL).						
FT	CARBOHYD	762	762	N-LINKED (GLCNAC. . .) (POTENTIAL).						
FT	CARBOHYD	762	762	N-LINKED (GLCNAC. . .) (POTENTIAL).						
SEQUENCE	774 AA;	87167 MW;	78CD1FB98A62424F CRC64;							
Query Match				35.8%;	Score 44;	DB 1;	Length 774;			
Best Local Similarity				41.2%;	Pred. No. 36;	Matches 7;	Conservative 3;</			





---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 16, 2001, 19:01:23 ; Search time 59.03 Seconds

(Without alignments)  
29,680 Million cell updates/sec

Title: US-09-675-650-3

Perfect score: 123

Sequence: 1 MFLHISPPKRYPHHTQEAQKEAQR 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	41.5	2492	1 A44213	nonstructural poly
2	51	41.5	2492	1 C44213	nonstructural poly
3	50	40.7	2492	1 MNMWD	nonstructural poly
4	46.5	37.8	357	2 T29856	probable aspartate
5	46.5	37.8	719	2 S55939	telomerase compone
6	46	37.4	602	2 T13219	major capsid prote
7	46	37.4	1138	2 A82939	membrane nuclease
8	45.5	37.0	349	2 D64825	hypothetical prote
9	45.5	37.0	349	2 F85598	hypothetical prote
10	45	36.6	673	2 B86437	hypothetical prote
11	44.5	36.2	419	2 T38024	MtIF protein - hum
12	44.5	36.2	511	1 VGVN	splee glycoprotein
13	44.5	36.2	520	2 T14752	microphthalmia-ass
14	44	35.8	91	2 E86163	protein P15K9.18 l
15	44	35.8	119	2 T00151	hypothetical prote
16	44	35.8	247	2 T17311	hypothetical prote
17	44	35.8	291	2 B86356	hypothetical prote
18	44	35.8	544	2 S44814	hypothetical prote
19	44	35.8	592	2 T25837	F4AB9.3 protein -
20	44	35.8	774	2 A39832	hypothetical prote
21	43.5	35.4	262	1 KOHU	scabrous locus (sc
22	43	35.0	186	2 B70040	tissue kallikrein
23	43	35.0	239	2 E81697	molybdenum transpo
24	43	35.0	250	2 H85067	ribose 5-phosphate
25	43	35.0	334	2 T23444	hypothetical prote
26	43	35.0	628	2 S01955	hypothetical prote
27	43	35.0	628	2 S19150	hypothetical prote
28	43	35.0	628	2 J00110	hypothetical 69K p
29	43	35.0	939	2 T05209	hypothetical prote

#### ALIGNMENTS

RESULT 1	43	35.0	1069	2 T00043	BH-protocadherin-a
A44213	43	35.0	1072	2 T00041	BH-protocadherin P
nonstructural polypeptide - Venezuelan equine encephalitis virus (strain P676)	32	43	1200	2 T00042	BH-protocadherin P
N:Contains: nonstructural protein NS1; nonstructural protein NS2; nonstructural prote	31	43	1200	2 T00042	nonstructural poly
C:Species: Venezuelan equine encephalitis virus	30	43	1200	2 S26372	nonstructural poly
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 17-Feb-1994	34	43	2493	2 S72349	microphthalmia-ass
C:Accession: A44213	35	42.5	419	2 A40728	microphthalmia-ass
R:Kinney, R.M.; Tsuchiya, K.R.; Snelder, J.M.; Trent, D.W.	36	42.5	442	2 T27676	hypothetical prote
Virology 191, 569-580, 1992	37	42.5	1003	2 T19638	hypothetical prote
A:Title: Genetic evidence that epizootic Venezuelan equine encephalitis (VEE) viruses	38	42	158	2 B85073	hypothetical prote
A:Reference number: A44213; MUID:93079859	39	42	201	1 E69503	indolepyruvate fer
A:Accession: A44213	40	42	341	2 A48936	nitrite reductase
A:Molecule type: genomic RNA	41	42	341	2 A82283	vibriolactin-type I
A:Residues: 1-2492 <KIN>	42	42	341	2 G81293	probable alpha-amy
A:Note: readthrough the terminator UGA occurs between the codons CAA for residue 1879	43	42	619	2 D71361	RNA polymerase II
C:Superfamily: Semliki Forest virus nonstructural protein	44	42	663	2 T49685	adenylate cyclase
C:Keywords: nonstructural protein; polypeptide	45	42	843	2 H82362	
F:1-535/Product: nonstructural protein NS1 #status predicted <NS1>					
F:586-1329/Product: nonstructural protein NS2 #status predicted <NS2>					
F:1330-1879/Product: nonstructural protein NS3 #status predicted <NS3>					
F:1880-2492/Product: nonstructural protein NS4 #status predicted <NS4>					
Query Match	41.5%	Score 51;	DB 1;	Length 2492;	
Best Local Similarity	33.3%	Pred. No. 21;			
Matches	7;	Conservative	8;	Mismatches	6;
				Indels	0;
				Gaps	0;
DB	1	MFLHISPPKRYPHHTQEAQKEA 21			
	1212	VFINWTPYKHYHOCEDHA 1232			
RESULT 2					
C44213					
nonstructural polypeptide - Venezuelan equine encephalitis virus (strain 3880)					
N:Contains: nonstructural protein NS1; nonstructural protein NS2; nonstructural prote					
C:Species: Venezuelan equine encephalitis virus					
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 17-Feb-1994					
C:Accession: C44213					
R:Kinney, R.M.; Tsuchiya, K.R.; Snelder, J.M.; Trent, D.W.					
Virology 191, 569-580, 1992					
A:Title: Genetic evidence that epizootic Venezuelan equine encephalitis (VEE) viruses					
A:Reference number: A44213; MUID:93079859					
A:Accession: C44213					
A:Molecule type: genomic RNA					
A:Residues: 1-2492 <KIN>					
A:Cross-references: GB:L00930					
A:Note: readthrough the terminator UGA occurs between the codons CAA for residue 1879					
C:Superfamily: Semliki Forest virus nonstructural protein					
C:Keywords: nonstructural protein; polypeptide					
F:1-535/Product: nonstructural protein NS1 #status predicted <NS1>					



DB 286 SSPSYSTYADAFK 299

## RESULT 7

A:Accession: A82939

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 &lt;STO&gt;

A:Cross-references: GB:AE005174; NID:g12513895; PIDN:AG55250.1; GSPDB:GN0145; UWGP:

A:Experimental source: strain 0157:H7, substrain EDL933

C:Genetics:

A:Gene: 21102

Query Match  
Best Local Similarity 37.0%; Score 45.5; DB 2; Length 349;  
Matches 9; Conservative 6; Mismatches 4; Indels 3; Gaps 1;

Query Match  
Best Local Similarity 45.0%; Score 46; DB 2; Length 1138;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 MFLHISPPKYPHTOAKE 20

DB 940 VFGHDSPTYSNPEYOKK 959

## RESULT 8

A:Accession: D64825

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1138 &lt;GTA&gt;

A:Cross-references: GB:AE002105; GB:AF222894; NID:g6899003; PIDN:AAF30460.1; GSPDB:GN001

A:Experimental source: serovar 3; biovar 1

C:Genetics:

A:Gene: mnuA-1; U0055

A:Genetic code: SGC3

Query Match  
Best Local Similarity 37.0%; Score 45.5; DB 2; Length 349;  
Matches 9; Conservative 6; Mismatches 4; Indels 3; Gaps 1;

QY 2 FLHISPPKYPHTOAKE 20

DB 121 FTHISSPSLYFDYHHRDIKED 142

## RESULT 9

A:Accession: F85598

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 &lt;STO&gt;

A:Cross-references: GB:AE000188; GB:U00096; NID:g1787084; PIDN:AACT73955.1; PID:g1787093;

A:Experimental source: strain K-12, substrain MG1655

C:Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase homology

Query Match  
Best Local Similarity 40.9%; Score 45.5; DB 2; Length 349;  
Matches 9; Conservative 6; Mismatches 4; Indels 3; Gaps 1;

Query Match  
Best Local Similarity 45.0%; Score 46; DB 2; Length 1138;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 FLHISPPKYPHTOAKE 20

DB 121 FTHISSPSLYFDYHHRDIKED 142

## RESULT 10

A:Accession: B86437

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-673 &lt;STO&gt;

A:Cross-references: GB:AE005172; NID:g4512629; PIDN:AA21698.1; GSPDB:GN00141

A:Map position: 1

Query Match  
Best Local Similarity 36.6%; Score 45; DB 2; Length 673;  
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 4 HISSPKYPHTOAKE 20

DB 339 HCFKPEWHLKANOE 355

## RESULT 11

A:Accession: I38024

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-419 &lt;RES&gt;

A:Cross-references: EMBL:Z29678; NID:g468496; PIDN:CAAB2775.1; PID:g468497

C:Genetics:

A:Gene: GDB:MTRF

A:Map position: 3p14.1-3p12

A:Map position: 3p14.1-3p12

A:Map position: 3p14.1-3p12

A:Map position: 3p14.1-3p12

A:Map position: 3p14.1-3p12

A:Map position: 3p14.1-3p12

A:Map position: 3p14.1-3p12

A:Map position: 3p14.1-3p12

A:Map position: 3p14.1-3p12

A:Map position: 3p14.1-3p12

A:Map position: 3p14.1-3p12

A:Map position: 3p14.1-3p12

A:Map position: 3p14.1-3p12

A:Map position: 3p14.1-3p12

A:Map position: 3p14.1-3p12

A:Map position: 3p14.1-3p12

A:Map position: 3p14.1-3p12

A:Map position: 3p14.1-3p12

Query Match 36.2%; Score 44.5; DB 2; Length 419;  
Best Local Similarity 47.1%; Pred. No. 33;  
Matches 8; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

OY 4 HISSPFKYPHTQEAQK 20  
| : | | | | : | : | : |  
DB 15 HLENPTKY-HIQQAQRQ 30

RESULT 12  
VGVN

spike glycoprotein G precursor - vesicular stomatitis Indiana virus (strain San Juan)  
C:Species: Vesicular stomatitis Indiana virus

C:Date: 31-Mar-1981 #sequence\_revision 19-Feb-1984 #text\_change 16-Jul-1999

C:Accession: A04117

R:Rose, J.K.; Gallione, C.J.

J. Virol. 39, 519-528, 1981

A:Title: Nucleotide sequences of the mRNA's encoding the vesicular stomatitis virus G ar

A:Reference number: A92983; MUID:82010868

A:Accession: A04117

A:Molecule type: mRNA

A:Residues: 1-511 <ROS>

A:Cross-references: GB:J02428; NID:9335873; PIDN:AAA48370.1; PID:9335877

C:Genetics:

A:Gene: G

C:Superfamily: rhabdovirus spike glycoprotein G

C:Keywords: glycoprotein; spike protein; transmembrane protein

F:1-22/Domian: signal sequence #status predicted <SIG>

F:23-511/Product: spike glycoprotein G #status predicted <SGG>

F:465-481/Domian: transmembrane #status predicted <TMN>

F:179,336/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 36.2%; Score 44.5; DB 1; Length 511;  
Best Local Similarity 42.9%; Pred. No. 40;

Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

OY 3 LHISSP---FKYPHTQEAQK 20  
| | | | | : | : | | | : | : | : |  
DB 412 LHSSKAQVFEPHPIQDAASQ 432

## RESULT 13

T14752

microphthalmia-associated transcription factor, MTF-A - human

C:Species: Homo sapiens (man)

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000

C:Accession: T14752; J0207

R:Kocher, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, August 1999

A:Reference number: Z18180

A:Accession: T14752

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-520 <KOE>

A:Cross-references: EMBL:AL110195

R:Name, S.; Fuse, N.; Yasumoto, K.; Sato, S.; Yajima, I.; Yamamoto, H.; Udono, T.K.; Dui

Biochem. Biophys. Res. Commun. 247, 710-715, 1998

A:Title: Identification of a novel isoform of microphthalmia-associated transcription fa

A:Reference number: J0207; MUID:98321192

A:Accession: J0207

A:Molecule type: mRNA

A:Residues: 1-520 <AMA>

A:Cross-references: DDBJ:AB006909; NID:93413846; PIDN:BA32286.1; PID:93413847

C:Comment: This protein plays a important role in differentiation of retinal pigment epi

C:Genetics:

A>Note: DKFZp586B2217.1

Query Match 36.2%; Score 44.5; DB 2; Length 520;

Best Local Similarity 47.1%; Pred. No. 41;  
Matches 8; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

OY 4 HISSPFKYPHTQEAQK 20  
| : | | | | : | : | : |  
DB 122 HLENPTKY-HIQQAQRQ 137

## RESULT 14

E86163

protein F15K9.18 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: E86163

R:Thellogis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Hulzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzla

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: E86163

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-91 <STO>

A:Cross-references: GB:AE05172; NID:93850586; PIDN:AMC72126.1; GSPDB:GN00141

C:Genetics:

A:Gene: F15K9.18

A:Map position: 1

Query Match 35.8%; Score 44; DB 2; Length 91;  
Best Local Similarity 41.7%; Pred. No. 7.9;

Matches 10; Conservative 3; Mismatches 7; Indels 4; Gaps 1;

OY 3 LHISSPFK---YPTHTQEAQK 22  
| | | | | : | | | | | : | : | : |  
DB 12 LHIRKVFEDVDPEKHTQNAKKEVE 35

## RESULT 15

T00151

hypothetical protein 18 - Staphylococcus aureus phage phi PVL

C:Species: Staphylococcus aureus phage phi PVL

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 21-Jul-2000

C:Accession: T00151

R:Kaneko, J.; Kimura, T.; Kawakami, Y.; Tomita, T.; Kamio, Y.

Biosci. Biotechnol. Biochem. 61, 1960-1962, 1997

A:Title: Pantone-Valentine leukocidin genes in a phage-like particle isolated from mlt

A:Reference number: Z14119; MUID:98067870

A:Accession: T00151

A:Status: translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-119 <KAN>

A:Cross-references: EMBL:AB009866; NID:93341907; PIDN:BA31891.1; PID:93341925

C:Superfamily: Staphylococcus aureus phage phi PVL hypothetical protein 18

Query Match 35.8%; Score 44; DB 2; Length 119;  
Best Local Similarity 47.1%; Pred. No. 10;  
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 2 FLHISSPFKYPHTQEAQ 18  
| : | | | | | : | : | : |  
DB 103 FAYFDGPKLRKPTDEVQ 119

Search completed: October 16, 2001, 19:04:50



Job time: 207 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 16, 2001, 18:30:52 ; Search time 53.97 Seconds  
(without alignments)  
8.775 Million cell updates/sec

Title: US-09-675-650-3  
Perfect score: 123  
Sequence: 1 MFHISPPKYPHTQEAQKRAQR 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfiles1.pep:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46.5	37.8	699	US-08-851-843A-52	Sequence 52, Appl
2	46.5	37.8	699	US-08-974-549A-188	Sequence 188, Appl
3	46.5	37.8	719	US-08-851-843A-7	Sequence 7, Appl
4	46.5	37.8	719	US-08-974-549A-219	Sequence 219, Appl
5	44.5	36.2	511	US-08-646-695-5	Sequence 5, Appl
6	44.5	36.2	511	PCR-US96-06053-5	Sequence 5, Appl
7	44	35.8	492	US-08-845-258-21	Sequence 21, Appl
8	44	35.8	492	US-08-990-571-21	Sequence 21, Appl
9	44	35.8	503	US-08-845-258-52	Sequence 52, Appl
10	44	35.8	503	US-08-990-571-52	Sequence 52, Appl
11	43.5	35.4	238	US-08-944-483-39	Sequence 39, Appl
12	43.5	35.4	258	US-08-744-026-3	Sequence 3, Appl
13	43.5	35.4	258	US-09-102-733-3	Sequence 3, Appl
14	43.5	35.4	258	US-09-261-767-3	Sequence 3, Appl
15	43.5	35.4	262	US-08-744-026-4	Sequence 4, Appl
16	43.5	35.4	262	US-08-790-137-1	Sequence 1, Appl
17	43.5	35.4	262	US-08-790-137-3	Sequence 3, Appl
18	43.5	35.4	262	US-08-681-151-4	Sequence 4, Appl
19	43.5	35.4	262	US-09-102-733-4	Sequence 4, Appl
20	43.5	35.4	262	US-08-824-874-4	Sequence 4, Appl
21	43.5	35.4	262	US-08-807-151-4	Sequence 4, Appl
22	43.5	35.4	262	US-09-261-767-4	Sequence 4, Appl
23	43.5	35.4	262	US-09-210-084-4	Sequence 4, Appl
24	43	35.0	747	US-09-035-648-18	Sequence 18, Appl
25	43	35.0	747	US-09-001-951-18	Sequence 18, Appl
26	42	34.1	127	US-08-853-659A-57	Sequence 57, Appl
27	41.5	33.7	432	US-08-522-166-8	Sequence 8, Appl

28	41.5	33.7	432	1	US-08-488-382A-8	Sequence 8, Appl
29	41.5	33.7	432	2	US-08-480-912-8	Sequence 8, Appl
30	40	32.5	301	5	PCT-US95-13975-72	Sequence 72, Appl
31	40	32.5	338	4	US-08-403-545-3	Sequence 3, Appl
32	40	32.5	338	4	US-08-404-381-3	Sequence 3, Appl
33	40	32.5	435	3	US-08-911-321-8	Sequence 2, Appl
34	39.5	32.1	1239	2	US-08-937-931-2	Sequence 2, Appl
35	39.5	32.1	1239	4	US-09-285-502-2	Sequence 2, Appl
36	39	31.7	40	3	US-08-926-842B-51	Sequence 51, Appl
37	39	31.7	313	3	US-08-926-842B-62	Sequence 62, Appl
38	38	30.9	488	1	US-07-794-393-2	Sequence 2, Appl
39	38	30.9	488	3	US-08-001-711-2	Sequence 2, Appl
40	38	30.9	488	1	US-08-704-711A-22	Sequence 22, Appl
41	38	30.9	489	4	US-08-448-488-11	Sequence 11, Appl
42	38	30.9	532	3	US-08-481-435-12	Sequence 12, Appl
43	38	30.9	553	3	US-08-481-435-11	Sequence 11, Appl
44	38	30.9	566	2	US-08-272-255-8	Sequence 8, Appl
45	38	30.9	566	5	PCT-US95-08565-8	Sequence 8, Appl

## ALIGNMENTS

RESULT 1  
US-08-851-843A-52  
Sequence 52, Application US/08851843A  
Patent No. 6093809  
GENERAL INFORMATION:  
APPLICANT: Cecch, Thomas R.  
APPLICANT: Lingner, Joachim  
APPLICANT: Nakamura, Toru  
APPLICANT: Chapman, Karen B.  
APPLICANT: Morin, Gregg B.  
APPLICANT: Harley, Calvin  
APPLICANT: Andrews, William H.  
TITLE OF INVENTION: No. 6093809e1 Telomerase  
NUMBER OF SEQUENCES: 225  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/851,843A  
FILING DATE: 06-MAY-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/724,643  
FILING DATE: 01-OCT-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002930US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 52:

SEQUENCE CHARACTERISTICS:  
LENGTH: 699 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-851-843A-52

Query Match 37.8%; Score 46.5; DB 3; Length 699;  
Best Local Similarity 41.4%; Pred. No. 9.7;  
Matches 12; Conservative 1; Mismatches 7; Indels 9; Gaps 1;

OY 2 FLHISP-----FKYPHTQEAQKEA 21  
| | | | | | | | | | | | | | | | | | | | | |  
DB 280 FCHISEPKERYKILGKKPKTEERYKAA 308

RESULT 2  
US-08-974-549A-188  
Sequence 188, Application US/08974549A  
Patent No. 6166178

GENERAL INFORMATION:  
APPLICANT: Cecch, Thomas R.  
APPLICANT: Lingner, Joachim  
APPLICANT: Nakamura, Toru  
APPLICANT: Chapman, Karen B.  
APPLICANT: Morlin, Gregg B.  
APPLICANT: Harley, Calvin B.  
APPLICANT: Andrews, William H.  
TITLE OF INVENTION: Human Telomerase Catalytic Subunit  
NUMBER OF SEQUENCES: 727  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/974,549A  
FILING DATE: 19-NOV-1997  
CLASSIFICATION: 536

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/724,643  
FILING DATE: 01-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/851,843  
FILING DATE: 06-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/854,050  
FILING DATE: 09-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/911,312  
FILING DATE: 14-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/912,951  
FILING DATE: 14-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/915,503  
FILING DATE: 14-AUG-1997  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US97/17618  
FILING DATE: 01-OCT-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US97/17885  
FILING DATE: 01-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph Ted  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002610US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 188:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 699 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-974-549A-188

Query Match 37.8%; Score 46.5; DB 4; Length 699;  
Best Local Similarity 41.4%; Pred. No. 9.7;  
Matches 12; Conservative 1; Mismatches 7; Indels 9; Gaps 1;

OY 2 FLHISP-----FKYPHTQEAQKEA 21  
| | | | | | | | | | | | | | | | | | | | | |  
DB 280 FCHISEPKERYKILGKKPKTEERYKAA 308

RESULT 3  
US-08-851-843A-7  
Sequence 7, Application US/08851843A  
Patent No. 6093809

GENERAL INFORMATION:  
APPLICANT: Cecch, Thomas R.  
APPLICANT: Lingner, Joachim  
APPLICANT: Nakamura, Toru  
APPLICANT: Chapman, Karen B.  
APPLICANT: Morlin, Gregg B.  
APPLICANT: Harley, Calvin B.  
APPLICANT: Andrews, William H.  
TITLE OF INVENTION: No. 6093809e1 Telomerase  
NUMBER OF SEQUENCES: 225  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94111

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/851,843A  
FILING DATE: 06-MAY-1997  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/724,643  
FILING DATE: 01-OCT-1996  
CLASSIFICATION:

APPLICATION NUMBER: US 08/911,312

```

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:

```

LENGTH: 511 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-646-695-5

Query Match 36.2%; Score 44.5; DB 4; Length 511;  
Best Local Similarity 42.9%; Pred. No. 14;  
Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

OY 3 LHSSP---FKYPHTEAOKE 20  
||:| | :| | | :| |  
Db 412 LHLSSKAQVEFHPIQDASQ 432

RESULT 6  
PCT-US96-06053-5  
Sequence 5, Application PC/TUS9606053  
GENERAL INFORMATION:  
APPLICANT: Yale University  
TITLE OF INVENTION: RECOMBINANT VESICULOVIRUSES AND THEIR  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/06053  
FILING DATE: 01-MAY-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: MISTOCK, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 6523-009-228  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 669-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 511 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-06053-5

Query Match 36.2%; Score 44.5; DB 5; Length 511;  
Best Local Similarity 42.9%; Pred. No. 14;  
Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

OY 3 LHSSP---FKYPHTEAOKE 20  
||:| | :| | | :| |  
Db 412 LHLSSKAQVEFHPIQDASQ 432

RESULT 7  
US-08-845-258-21  
Sequence 21, Application US/08845258  
Patent No. 6183976  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
ATTORNEY/AGENT INFORMATION:  
NAME: Lodes, Michael J.

APPLICANT: Houghton, Raymond  
APPLICANT: Sleath, Paul R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: SEED AND BERRY  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/845,258  
FILING DATE: 24-APR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MAKI, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.426C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 492 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-845-258-21

Query Match 35.8%; Score 44; DB 4; Length 492;  
Best Local Similarity 31.8%; Pred. No. 17;  
Matches 7; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

OY 2 FLHSSPKYPTHTAOKEAQR 23  
| :| | :| | | :| |  
Db 103 FVYISKEYEHTELAKCHK 124

RESULT 8  
US-08-990-571-21  
Sequence 21, Application US/08990571  
Patent No. 6214971  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G. et al.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B.  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: SEED AND BERRY  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/990,571  
FILING DATE: 11-DEC-1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: MAKI, David J.  
REGISTRATION NUMBER: 31,392

APPLICANT: COHEN, MAURICE  
APPLICANT: COLPITS, TRACEY L.  
APPLICANT: FRIDMAN, PAULA N.  
APPLICANT: GRAMADOS, EDWARD N.  
APPLICANT: KLASS, MICHAEL R.  
APPLICANT: RUSSELL, JOHN C.  
APPLICANT: STEWART, KENT D.  
APPLICANT: STROUPE, STEVEN D.  
TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS  
AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES  
OF THE PROSTATE

```

NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Fastseq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,483
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Becker, Cheryl L.
REGISTRATION NUMBER: 35,441
REFERENCE/DOCKET NUMBER: 6183. US. 01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847/935-1729
TELEFAX: 847/938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: NO. 6232456e
OS-08-944-483-39

```

```

? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Billings, Lucy J
? REGISTRATION NUMBER: 36,749
? REFERENCE/DOCKET NUMBER: PF-0154 USA
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415-855-0555
? TELEFAX: 415-845-4166
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 258 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? IMMEDIATE SOURCE:
? LIBRARY: GenBank
? CLONE: 871814
? US-08-744-026-3

```





---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2001, 17:49:27 ; Search time 1981.17 Seconds  
(without alignments)  
2414.303 Million cell updates/sec

Title: US-09-675-650-1

Perfect score: 506

Sequence: 1 cagggaagcacaagaagc.....ggcttcacaagacatgcaac 506

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_esthum29:\*  
64: em\_esthum30:\*  
65: em\_esthum31:\*  
66: em\_esthum32:\*  
67: em\_esthum33:\*  
68: em\_esthum34:\*  
69: em\_esthum35:\*  
70: em\_esthum36:\*  
71: em\_esthum37:\*  
72: em\_esthum38:\*  
73: em\_esthum39:\*  
74: em\_esthum40:\*  
75: em\_esthum41:\*  
76: em\_esthum42:\*  
77: em\_esthum43:\*  
78: em\_esthum44:\*  
79: em\_esthum45:\*  
80: em\_esthum46:\*  
81: em\_esthum47:\*  
82: em\_esthum48:\*  
83: em\_esthum49:\*  
84: em\_esthum50:\*  
85: em\_esthum51:\*  
86: em\_esthum52:\*  
87: em\_esthum53:\*  
88: em\_esthum54:\*  
89: em\_esthum55:\*  
90: em\_esthum56:\*  
91: em\_esthum57:\*  
92: em\_esthum58:\*  
93: em\_esthum59:\*  
94: em\_esthum60:\*  
95: em\_esthum61:\*  
96: em\_esthum62:\*  
97: em\_esthum63:\*  
98: em\_esthum64:\*  
99: em\_esthum65:\*  
100: em\_esthum66:\*  
101: em\_esthum67:\*  
102: em\_esthum68:\*  
103: em\_esthum69:\*  
104: em\_esthum70:\*  
105: em\_esthum71:\*  
106: em\_esthum72:\*  
107: em\_esthum73:\*  
108: em\_esthum74:\*  
109: em\_esthum75:\*  
110: em\_esthum76:\*  
111: em\_esthum77:\*  
112: em\_esthum78:\*  
113: em\_esthum79:\*  
114: em\_esthum80:\*  
115: em\_esthum81:\*  
116: em\_esthum82:\*

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: em_gss_fun:*
182: em_gss_hum1:*
183: em_gss_hum2:*
184: em_gss_hum3:*
185: em_gss_hum4:*
186: em_gss_hum5:*
187: em_gss_hum6:*
188: em_gss_hum7:*
189: em_gss_hum8:*
190: em_gss_hum9:*
191: em_gss_inv1:*
192: em_gss_inv2:*
193: em_gss_inv3:*
194: em_gss_inv4:*
195: em_gss_inv5:*
196: em_gss_inv6:*
197: em_gss_inv7:*
198: em_gss_inv8:*
199: em_gss_inv9:*
200: em_gss_inv10:*
201: em_gss_inv11:*
202: em_gss_inv12:*
203: em_gss_inv13:*
204: em_gss_inv14:*
205: em_gss_inv15:*
206: em_gss_inv16:*
207: em_gss_inv17:*
208: em_gss_inv18:*
209: em_gss_inv19:*
210: em_gss_inv20:*
211: em_gss_inv21:*
212: em_gss_inv22:*
213: em_gss_inv23:*
214: em_gss_inv24:*
215: em_gss_inv25:*
216: em_gss_inv26:*
217: em_gss_inv27:*
218: em_gss_inv28:*
219: em_gss_inv29:*
220: em_gss_inv30:*
221: em_gss_inv31:*
222: em_gss_inv32:*
223: em_gss_inv33:*
224: em_gss_inv34:*
225: em_gss_inv35:*
226: em_gss_inv36:*
227: em_gss_inv37:*
228: em_gss_inv38:*
229: em_gss_inv39:*
230: em_gss_inv40:*
231: em_gss_inv41:*
232: em_gss_inv42:*
233: em_gss_inv43:*
234: em_gss_inv44:*
235: em_gss_inv45:*
236: em_gss_inv46:*
237: em_gss_inv47:*
238: em_gss_inv48:*
239: em_gss_inv49:*
240: em_gss_inv50:*
241: em_gss_inv51:*
242: em_gss_inv52:*
243: em_gss_inv53:*
244: em_gss_inv54:*
245: em_gss_inv55:*
246: em_gss_inv56:*
247: em_gss_inv57:*
248: em_gss_inv58:*
249: em_gss_inv59:*
250: em_gss_inv60:*
251: em_gss_inv61:*
252: em_gss_inv62:*
253: em_gss_inv63:*
254: em_gss_inv64:*
255: em_gss_inv65:*
256: em_gss_inv66:*
257: em_gss_inv67:*
258: em_gss_inv68:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

[illegible]











[illegible][illegible]

[illegible]

FEATURES	Source	Location/Qualifiers
BASE COUNT	176 a 169 c 124 g 214 t 348 others.	
ORIGIN	<pre> 1. 1031 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPC1-98" /clone="BACR25K05" /note="end : TET3" </pre>	
QUERY MATCH	7.7%; Score 38.8; DB 219; Length 1031;	
Best Local Similarity	25.3%; Pred. No. 1.6;	
Matches	76; Conservative 85; Mismatches 139; Indels 0; Gaps 0;	
QY	<pre> 11 aaagggaagccagaggaagtgctttataagaacccaattctactcagaattttg 70    ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::    989 AAAAAAAAAAAMAMAMNACAAAMAMAMAMAMATCTCTCTMTMYATATMYNYNC 930 </pre>	
QY	<pre> 71 atggccttaagtcctctaccgcgttctctatccctccctaccgcgtcccggaatca 130 ::: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::  929 HTTAHTATAMTYHMACAGATYTTMTWTWTWCMTACTCTCYCHTHTYHTYHMAHTATMA 870 </pre>	
QY	<pre> 131 ctaccagattctatcttctcgtctcgatgtctgactgactgcattgattatccctca 190    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::  869 CTMCCHMYMYMHYMTCTAHTATCTMCCTCYMTWTWYHCCCTTAHTYTWHTHTYTTWT 810 </pre>	
QY	<pre> 191 cggagctcggatttctcaaccggcgctcaaccctccctccatattgtctccacttc 250    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::  809 AAMNMTTYMACMCTMAHCYCHMAAAYMAAYMYTWYCYMWCSCCAATWTWTTCTMMNACTCTM 750 </pre>	
QY	<pre> 251 acagatccctcgtggagaatcccgccgcgcacatcttg99tcatgatgagctcgccctgt 310 :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::  749 HCCMWCACCTMYMYTCYTHMYTCYTCCTCCYCMYTYTYAHTMAHMAHMAACCAATACNTAT 690 </pre>	
RESULT 15	AO895392 1049 bp DNA GSS 10-NOV-1999	
LOCUS	HS-3143-B2.C10-T7C CIT Approved Human Genomic Sperm Library D Homo	
DEFINITION	sapliens genomic clone Plate=3143 COL=20 Kow=F, DNA sequence.	
ACCESSION	AO895392	
VERSION	AO895392.1 GI:6351582	
KEYWORDS	GSS.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
	1 (bases 1 to 1049)	
	Mahaitas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,	
	Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and	
	Hood,L.	
TITLE	Sequence-tagged connectors: A sequence approach to mapping and	
JOURNAL	scanning the human genome	
MEDLINE	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)	
COMMENT	99380589	
	Contact: Mahaitas GG, Wallace JC, Hood L	
	High Throughput Sequencing Center	
	University of Washington	

401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com)  
BMC end Web Server: <http://www.htsc.washington.edu>  
Plate: 3143 row: F column: 20

Class: BAC ends  
High quality sequence stop: 1049.

## FEATURES

**Source**

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1bp="CIT Approved Human Genomic Sperm Library D"
/clone_1bp="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/notes="Organ: sperm; Vector: pBelBac11; BAC Clones in E-Coli DH10B"
BASE COUNT      347 a      572 g      55 t      43 others
ORIGIN

```

Query Match 7.7%; Score 38.8; DB 235; Length 1049;

Best Local Similarity 51.18; Pred. No. 1.6;  
Matches 01; Conservative 0; Mismatches

Matches	91;	Conservative	0;	Mismatches	87;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

83 tccttactcgttctatccttcctactcactgttcctccggaatccactacgatttc 142

Db 771 TACCTCTCTTTTCTCTCTTCCACCCCTCCCTCTCTCTCTCCCTCATTCAC 712

QY 143 cttctctgcctgtattgtctgactggtcactgtgattatcctcagcagtgat 202

Db 711 TTTCTCCTTCTCCTCCTCCCTCCTCCCTTCCTTCCCTCCCTTCCTCTCTT 652

QY 203 ttctaccgggtcacctcgcctccataattgtcctccacttcacagatccct 260

Db 651 TCTCCTTCTCCTCCCTCCCTTCTCCTCCCTCCCTTCCCTCCTCTCTCTCT 594

Search completed: October 16, 2001, 18:30:40  
Job time: 2473 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2001, 17:53:02 ; Search time 88.04 Seconds  
(Without alignments) 1088.044 Million cell updates/sec

Title: US-09-675-650-1  
Perfect score: 506  
Sequence: 1 ccggaagcacaagaagc.....ggttcacaagacatgcaac 506

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PCYUS.COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58.2	11.5	7218	1	US-08-232-463-14 Sequence 14, Appl
2	35	6.9	10207	1	US-08-920-812-2 Sequence 2, Appl
3	35	6.9	10207	1	US-08-920-812-2 Sequence 2, Appl
4	35	6.9	10207	1	US-08-921-177-2 Sequence 2, Appl
5	35	6.9	10207	1	US-08-362-577C-2 Sequence 2, Appl
6	35	6.9	10207	1	US-08-920-828-2 Sequence 2, Appl
7	33.2	6.6	152331	3	US-09-128-155-16 Sequence 16, Appl
8	33	6.5	7218	1	US-08-232-463-14 Sequence 14, Appl
9	32.6	6.4	46899	1	US-08-471-119A-1 Sequence 17, Appl
10	31.6	6.2	176373	3	US-09-128-155-17 Sequence 17, Appl
11	30.6	6.0	4758	3	US-09-191-647-1 Sequence 1, Appl
12	29.8	5.9	53526	3	US-08-658-136-1 Sequence 2, Appl
13	29.8	5.9	53577	3	US-08-658-136-1 Sequence 2, Appl
14	29.6	5.8	4177	3	US-09-023-082A-23 Sequence 23, Appl
15	29.4	5.8	2085	2	US-08-668-128B-7 Sequence 7, Appl
16	29.4	5.8	2085	2	US-08-905-445-7 Sequence 7, Appl
17	29.2	5.8	1838	5	PCT-US93-06251-85 Sequence 85, Appl
18	29.2	5.8	4175	1	US-08-306-691B-49 Sequence 49, Appl
19	29.2	5.8	4175	4	US-08-202-841A-1 Sequence 1, Appl
20	29.2	5.8	4175	5	PCT-US93-06251-84 Sequence 84, Appl
21	29.2	5.8	7785	2	US-08-276-967-1 Sequence 1, Appl
22	29.2	5.8	152331	3	US-09-128-155-16 Sequence 16, Appl
23	29	5.7	4517	4	US-09-140-804-9 Sequence 9, Appl
24	28.8	5.7	1022	4	US-08-960-780-33 Sequence 33, Appl
25	28.8	5.7	1022	4	US-09-073-898-33 Sequence 1, Appl
26	28.8	5.7	1952	1	US-08-333-358-1 Sequence 1, Appl
27	28.8	5.7	1952	1	US-08-463-694-1 Sequence 1, Appl

c	28	28.8	5.7	1952	1	US-08-694-501-1	Sequence 1, Appl
c	29	28.8	5.7	3460	2	US-08-751-305-1	Sequence 1, Appl
c	30	28.8	5.7	4074	1	US-08-471-033-19	Sequence 19, Appl
c	31	28.8	5.7	4074	2	US-08-471-044-19	Sequence 19, Appl
c	32	28.8	5.7	4074	2	US-08-463-483A-19	Sequence 19, Appl
c	33	28.8	5.7	4074	2	US-08-471-046A-19	Sequence 19, Appl
c	34	28.8	5.7	4074	2	US-08-470-566B-19	Sequence 19, Appl
c	35	28.8	5.7	4074	2	US-08-469-334-19	Sequence 19, Appl
c	36	28.8	5.7	4074	3	US-09-300-529-19	Sequence 19, Appl
c	37	28.8	5.7	9370	1	US-08-320-559-27	Sequence 27, Appl
c	38	28.8	5.7	9370	3	PCT-US94-04496-27	Sequence 27, Appl
c	39	28.8	5.7	9370	5	PCT-US94-04496-27	Sequence 27, Appl
c	40	28.8	5.7	9391	1	US-08-320-559-25	Sequence 25, Appl
c	41	28.8	5.7	9391	3	US-08-545-860D-25	Sequence 25, Appl
c	42	28.8	5.7	9391	5	PCT-US94-04496-25	Sequence 25, Appl
c	43	28.6	5.7	301	2	US-08-332-766A-23	Sequence 23, Appl
c	44	28.4	5.6	291	1	US-07-922-723A-7	Sequence 7, Appl
c	45	28.4	5.6	291	1	US-07-799-828C-7	Sequence 7, Appl

## ALIGNMENTS

RESULT 1  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOXLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-Fls  
US-08-232-463-14

[illegible]

RESULT 2  
 US-08-920-812-2/c  
 Sequence 2, Application US/08920812  
 Patent No. 5763188  
 GENERAL INFORMATION:  
 APPLICANT: Ohno, Tsuneya  
 APPLICANT: Matsuhisa, Akio  
 APPLICANT: Uehara, Hirotsugu  
 APPLICANT: Eda, Soji  
 TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
 NUMBER OF SEQUENCES: 25  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 STREET: 6300 Sears Tower, 233 South Wacker Drive  
 City: Chicago  
 STATE: Illinois  
 COUNTRY: United States of America  
 Zip: 60606-6402  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/920,812  
 FILING DATE: 29-AUG-1997  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/362,577  
 FILING DATE: 27-MAR-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Rin-laures, Li-Hsien  
 REGISTRATION NUMBER: 33,547  
 REFERENCE/DOCKET NUMBER: 19036/32420  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/474-6300  
 TELEFAX: 312/474-0448  
 TELEEX: 25-3856  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 10207 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA  
 ORIGINAL SOURCE:  
 ORGANISM: Staphylococcus aureus  
 STRAIN: Clinical Isolate SA-24

Query Match	Best Local Similarity	Score	DB 1	Length
Matches 101; Conservative	47.9%;	6.9%;	35;	10207;
			0; Mismatches 110;	Indels 0; Gaps 0;
Qy 283	ctgttgatcatgatgaagcctgcctgtgcctgtgcttgcgcctgtggaaggaagacattag	342		
Db 2910	CATCACTCTTTCATTAAGCGTAGCATTCGATTCACCTCGCATTTGATGTGAATAACTTCAA	2851		
Qy 343	aaaatgaattgtagtgccttccctaagaatggygcagaaacagatccctgttggatatt	402		
Db 2850	TCATTAAGCAAGATTCATTTCCATGACCGGATGATGTCATGCAACGCTGTATCAATGT	2791		
Qy 403	tatttgaacggattacagatttggaaatgaagtcaaccaagtggacattccaatgagag	462		
Db 2790	TATTAAGCAATATTATTAAACCGGTAAATATGCAATCAAAATTAATAAGCATTAAGACTCGTCAC	2731		
Qy 463	gaaaacagacgagaatatttgatgcttca	493		
Db 2730	GATTAATAAATTCGAAATCTTGATTCGTTAA	2700		

RESULT 3  
 US-08-920-827-2/c  
 Sequence 2, Application US/08920827  
 Patent No. 570375  
 GENERAL INFORMATION:  
 APPLICANT: Ohno, Tsuneya  
 APPLICANT: Matsuhisa, Akio  
 APPLICANT: Uehara, Hirotsugu  
 APPLICANT: Eda, Soji  
 TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
 NUMBER OF SEQUENCES: 25  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 STREET: 6300 Sears Tower, 233 South Wacker Drive  
 CITY: Chicago  
 STATE: Illinois  
 COUNTRY: United States of America  
 ZIP: 60606-6402  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/920,827  
 FILING DATE: 29-AUG-1997  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/362,577  
 FILING DATE: 27-MAR-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Rijnlaars, Li-Hsien  
 REGISTRATION NUMBER: 33,347  
 REFERENCE/DOCKET NUMBER: 19036/32420  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/474-6300  
 TELEFAX: 312/474-0448  
 TELEX: 25-3856  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 10207 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA  
 ORIGINAL SOURCE:  
 ORGANISM: *Staphylococcus aureus*  
 STRAIN: Clinical Isolate SA-24  
 US-08-920-827-2

	Query Match	6.9% ; Score 35 ; DB 1 ; Length 10207 ;
	Best Local Similarity	47.9% ; Pred. No. 0.2 ;
	Matches 101 ; Conservative 0 ; Mismatches 110 ; Indels 0 ; Gaps 0 ;	
OY	283 ctgagtcacatcgaatgagccctgcgccctgtgcctgtgtcccgtctgtgaggaaagacattag	342
Db	2910 CATGACGTTCATTAAGGGTGAAGCATTCGATTGCATTCACATTCGCCAATTATGAAATAACTCA	2851
OY	343 aaactgaattgatctgttcccttaagaagtgtggcaagaaaacagcatcccgltgtgataatt	402
Db	2850 TCAATGAACGCAATCATTCCATGACC GGATGTGTGTCTATCCAAGCCTTTGTATCAANGT	2791
OY	403 tatttgaacgagattacagatttgaattgaagtcaaccaaatgtgacattaccaatlgagag	462
Db	2790 TATTAAGCAATATTATAAGCGGTATATATGGAATCAATAATAAACATTAAGACTGTAC	2731
OY	463 gaaaacagacagaanaatcttgatggctcca	493
Db	2730 GAATATAAATTCATATCTTGATATTCGTAA	2700

```

RESULT 4
US-08-921-177-2/c
: Sequence 2, Application US/08921177
Patent No. 5798211
: GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsubisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Diseases
NUMBER OF SEQUENCE: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/921,177
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 10207 base pairs
type: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-24
US-08-921-177-2

```

	Query Match	Best Local Similarity	Score 35;	DB 1;	Length 10207;
			6.9%;		
			Pred. No. 0.2;		
	Matches 101;	Conservative 0;	Mismatches 110;	Indels 0;	Gaps 0;
QY	283	cttggcgcacgatgtagcctcgccctgtgcctgtgiccgcctgttgagaggaagacattag	342		
DB	2910	CATCAGCTTTCAATTAAGGATGAGCATTTGGATTCCATTCGCAATTAATTAAGTAACCTTCA	2851		
QY	343	aaatgaattgatgtgcttccttaagagatggcagaaacagatccctgttgatatt	402		
DB	2850	TCAATGAACGAAATCTTTCCATGACCGGATGATGTGTCAATCAACGCTTTGTATCAATCT	2791		
QY	403	tattgaacgagattacagatttgaatatgaagttccaaagtgaagattaccaatgagag	462		
DB	2790	TATAAGGAAATATTATAAAGCGGTATATGGAATCAATAATAAAGCAATAAGACTGTCAC	2731		
QY	463	gaaacagacagaaatcttgatgcttca	493		
DB	2730	GATATAAATTGCATATCTTGATATCCGTA	2700		

```

RESULT      5
US-08-362-577C-2/c
Sequence 2, Application US/08362577C
Patent No. 5807673
GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,577C
FILING DATE: 27-MAR-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 10207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-24
US-08-362-577C-2
Query Match      6.9%; Score 35; DB 1; Length 10207;
Best Local Similarity 47.9%; Pred. No. 0.2; Indels 0; Gaps 0;
Matches 101; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

```

QY 283 cttggtgcacgtctgtagccctcgcccgctgcctcggtcgtgtgagggaagacattag 342  
 Db 2910 CATCAGCTTCAATAAAGCGATGATTGATCTCATCTGGCCAAATTGAATTGAATACCTTCAA 2851  
 QY 343 aaaaatgatgatgctgtcccttaagaagctggcgaaagacagatccctgttgtagatc 402  
 Db 2850 TCATGAAAGCAATTCATTTCCATCAGCGGATGATGTCTCATCAAGCGTTTGTATCAATGT 2791  
 QY 403 taattgaagcggaattacaagaattgaagtgaaagtcacaaagtgtagcattaccatlagag 462  
 Db 2790 TATAGCAAAATTATTAAGGCGTAATATGGAATCAAAATTAATACATTAAGCACTGCTCAC 2731  
 QY 463 gaaacacgacgaaaaactctgttgctctca 493  
 Db 2730 GATTATAAATTGCAATATCTTGATATATCTAA 2700

```

RESULT      6
US-08-920-828-2/c
: Sequence 2, Application US/08920828
: Patent No. 5853998
:
: GENERAL INFORMATION:
:   APPLICANT: Ohno, Tsuneya
:   APPLICANT: Matsubae, Aki
:   APPLICANT: Uehara, Hirotsugu
:   TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
:   NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESS:
:   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
:   STREET: 6300 Sears Tower, 233 South Wacker Drive
:   CITY: Chicago
:   STATE: Illinois
:   COUNTRY: United States of America
:   ZIP: 60606-6402
:
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: PatentIn Release #1.0, Version #1.25
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/920,828
:     FILING DATE: 29-AUG-1997
:     CLASSIFICATION: 435
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: US 08/362,577
:       FILING DATE: 27-MAR-1995
:     ATTORNEY/AGENT INFORMATION:
:       NAME: Rih-Laures, Li-Hsien
:       REGISTRATION NUMBER: 33,547
:       REFERENCE/DOCKET NUMBER: 19036/32420
:     TELECOMMUNICATION INFORMATION:
:       TELEPHONE: 312/474-6300
:       TELEFAX: 312/474-0448
:       TELEX: 25-3856
:     INFORMATION FOR SEQ ID NO: 2:
:       SEQUENCE CHARACTERISTICS:
:         LENGTH: 10207 base pairs
:         TYPE: nucleic acid
:         STRANDEDNESS: double
:         TOPOLOGY: linear
:       MOLECULE TYPE: Genomic DNA
:       ORIGINAL SOURCE:
:         ORGANISM: Staphylococcus aureus
:         STRAIN: Clinical Isolate SA-24
:
: US-08-920-828-2

```

Query Match	6.98;	Score 35;	DB 2;	Length 10207;
Best Local Similarity	47.98;	Pred. NO. 0.2;		
Matches 101;	Conservative	0;	Mismatches 110;	Indels 0;
			Gaps	0;

Oy	283	cttgggtcatcagatgaagccctgcgcctctgcctgcctctctgtaaggaagaacattag	342
Db	2910	CATAGCTTTCATATAACCGTAGCATTTGGATTCACTCTCCCAATTAGTTSAAATACCTTCA	2851
Oy	343	aaatgtaatgatgtgtctcccttaaagatgtggcaggaaaacagatcctgttgyatatc	402
Db	2850	TCATGGAACGATCATTTTCATGACCGGATGATGTCTCATCGAACGCTTTGTATCAATGT	2781
Oy	403	tatttgaacgggattacagatttgaatgaagtacccaagtfgagcatctacaaatgagaag	462
Db	2790	TATTAAGCAATATTATTAAGCGGTAAATATGAAACAATTAATTAAGCATTAAGACTGCTGAC	2731
Oy	463	gaaacagacagagaatacttga tggcttca	493
Db	2730	GATTAATAATTGCATATCTTGTATACGTAA	2700

```

RESULT 7
US-09-128-155-16/C
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

```

	Query Match	6.6%	Score 33.2	DB 3	Length 152331
	Best Local Similarity	51.3%	Prod. No. 3.7		
	Matches 77	Conservative	1	Mismatches 73	Indels 0
				Gaps 0	
Oy	82	ttccctactcgttctcttcctctactcaatgctcccggaatccactacgatatt	141		
Db 147068		ttcccttccctcgttcccttctgcttcttcttcttcccttcccttcttcttcccttctt	147069		
Oy	142	ctattctctgcctglatgtctgactgcacttgatattatcccaaggagctcga	201		
Db 147008		tctcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt	146949		
Oy	202	ttttctaccggggtcactcctcgtctccccc	231		
Db 146948		tcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt	146919		

RESULT 3 8  
 US-08-232-463-14/C  
 : Sequence 14, Application US/08232463  
 : Patent No. 5670367  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: DORNER, F.  
 : APPLICANT: SCHUEFLINGER, F.  
 : APPLICANT: FALKNER, F. G.  
 : TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
 : NUMBER OF SEQUENCES: 52  
 : CORRESPONDENCE ADDRESS:  
 :



```

; ADDRESSER: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F15
; US-08-232-463-14

Query Match 6.5%; Score 33; DB 1; Length 7218;
Best Local Similarity 3.5%; Pred. NO. 0.75;
Matches 6; Conservative 105; Mismatches 60; Indels 0; Gaps 0;

QY 327 gaggaagacattagaataatgattgctgtctcttaagatggcgagaaacaga 366
DB 1229 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1170
QY 387 tccgtgtgatatattatgtaacggaattacagattgaagtcaccaaagtg 446
DB 1169 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1110
QY 447 gcattccaatgaggaagaaacagacgagaaactctgattgctcaaga 497
DB 1109 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1059

RESULT 9
US-08-471-119A-1
; Sequence 1, Application US/08471119A
; Patent No. 5827706
; GENERAL INFORMATION:
; APPLICANT: Leitner, Ernst
; APPLICANT: Schneider, Elisabeth
; APPLICANT: Schoergerdorfer, Kurt
; APPLICANT: Weber, Gerhard
; TITLE OF INVENTION: Cyclosporin Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: No. 5827706artis Corporation
; STREET: 59 Route 10
; CITY: East Hanover
; STATE: New Jersey
; COUNTRY: USA
```

```

; ZIP: 07936
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,119A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kassenoff, Melvyn
; REGISTRATION NUMBER: 26,389
; REFERENCE/DOCKET NUMBER: 100-8029/CONF/CONF
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 503 8474
; TELEFAX: 201 503 8807
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46899 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Tolypocladium niveum
; STRAIN: ATCC 34921
; US-08-471-119A-1

Query Match 6.4%; Score 32.6; DB 1; Length 46899;
Best Local Similarity 60.9%; Pred. No. 3;
Matches 53; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 185 tctcaacgagctgctgattctaccgaggtcacctcgctccctcatattgctc 244
DB 38506 TTCTGCCCGGCTCTAGAGATGCTCCCGGATATCCCTGTTCAACATCATTTGTGCTCC 38565
QY 245 attccacagatccctggagaaatgc 271
DB 38566 ACTCCAGCAGAAATCTCGTAACTGC 38592

RESULT 10
US-09-128-155-17
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; EARLIER FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(176373)
; OTHER INFORMATION: n = A,T,C or G
; US-09-128-155-17
```

```

Query Match          6.2%: Score 31.6; DB 3; Length 176373;
Best Local Similarity 50.7%: Pred. No. 14;
Matches 76; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

OY      82  ttctctactcgtttctctactcctctactcactgtctctcccggaatccactacgatttt 141
        |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 122325  ttcttctcgtctgctgtcgtctcttctcttctctctctctctctcttctctcttct 122384

OY      142  ctatttctgtccctcgatctgtctgacgtgcctcacttgattatctctcaggaactgga 201
        ||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 122385  tctcttctctctctctctctctctctctctctctctctctctctctctctctctt 122444

OY      202  ttctctaccgggtccactccgtccgtccccc 231
        ||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 122445  tctacttcttctctctctctctctctctctc 122474

RESULT 11
US-09-191-647-1
: Sequence 1, Application US/09191647
: Patent No. 6046015
: GENERAL INFORMATION:
: APPLICANT: Goodman, Corey
: APPLICANT: Kid, Thomas
: APPLICANT: Brosse, Katja
: APPLICANT: Tessier-Lavigne, Marc
: TITLE OF INVENTION: Modulating Robo: Ligand Interactions
: FILE REFERENCE: B98-031-3
: CURRENT APPLICATION NUMBER: US/09/191, 647
: CURRENT FILING DATE: 1998-11-13
: EARLIER APPLICATION NUMBER: 60/065,544
: EARLIER FILING DATE: 1997-11-14
: EARLIER APPLICATION NUMBER: 60/081,057
: EARLIER FILING DATE: 1998-04-07
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 4758
: TYPE: DNA
: ORGANISM: human
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(4575)
US-09-191-647-1

Query Match          6.0%: Score 30.6; DB 3; Length 4758;
Best Local Similarity 53.8%: Pred. No. 3.7;
Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY      279  ccaatcttggtgcatcgaatgcacgtccctctgtccttgtlcccgctgtgtgaggaaggaca 338
        |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2824  ccaatgtgttcgaaggcgacgacgtgtggtcccaatcatcgatcgatcaaccaca 2883

OY      339  ttgaagaatgaatgatgtgtgtctcttaagaatggtgcaggaataacagatccctgtgt 395
        ||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2884  tgaatacatggaaggaactgtccacttaagaagaagaagaagaatgatctctgtgt 2940

RESULT 12
US-08-658-136-2
: Sequence 2, Application US/08658136
: Patent No. 6071717
: GENERAL INFORMATION:
: APPLICANT: KLINGER, KATHERINE W
: APPLICANT: LANDES, GREGORY M
: APPLICANT: BURN, TIMOTHY C
: APPLICANT: CONNORS, TIMOTHY D
: APPLICANT: DACKOWSKI, WILLIAM
: APPLICANT: GERMINO, GREGORY
: APPLICANT: QIAN, FENG
: TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
: NUMBER OF SEQUENCES: 58

```

```

1 CORRESPONDENCE ADDRESS:
2 ADDRESSEE: GENZYME CORPORATION
3 STREET: ONE MOUNTAIN ROAD
4 CITY: FRAMINGHAM
5 STATE: MASSACHUSETTS
6 COUNTRY: USA
7 ZIP: 01701
8
9 COMPUTER READABLE FORM:
10 MEDIUM TYPE: Floppy disk
11
12 ATTORNEY/AGENT INFORMATION:
13 NAME: LASSEN, ELIZABETH
14 REGISTRATION NUMBER: 31,845
15 REFERENCE/DOCKET NUMBER: GEN4-17,8
16 TELECOMMUNICATION INFORMATION:
17 TELEPHONE: 508-872-8400
18 TELEFAX: 508-872-5415
19
20 INFORMATION FOR SEQ ID NO: 2:
21 SEQUENCE CHARACTERISTICS:
22 LENGTH: 53526 base pairs
23 TYPE: nucleic acid
24 STRANDEDNESS: single
25 TOPOLOGY: linear
26
27 MOLECULE TYPE: DNA (genomic)
28
29 US-08-658-136-2
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```

[illegible]



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2001, 17:56:37 ; Search time 164.45 Seconds  
(without alignments)  
1932.005 Million cell updates/sec

Title: US-09-675-650-1

Perfect score: 506  
Sequence: 1 caggaagcacaagaaggaac.....ggcttcacagaacatgcac 506

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues  
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 405 summaries

Database : N\_Geneseq.0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT:\*  
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT:\*  
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT:\*  
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT:\*  
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT:\*  
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT:\*  
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT:\*  
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT:\*  
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT:\*  
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT:\*  
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT:\*  
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT:\*  
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT:\*  
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT:\*  
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT:\*  
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT:\*  
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT:\*  
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT:\*  
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT:\*  
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT:\*  
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT:\*  
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	243.4	48.1	812	21	Human immunogenic
2	243.4	48.1	820	19	Prostate cancer an
3	243.4	48.1	1872	19	Prostate cancer an
4	243.4	48.1	2037	19	Prostate cancer an
5	243.4	48.1	2229	21	Human immunogenic
6	243.4	48.1	2426	21	Human immunogenic
7	243.4	48.1	3112	21	Human immunogenic
8	243.4	48.1	3582	19	Prostate cancer an
9	224.6	44.4	597	20	Human secreted pro
10	218.4	43.2	718	21	Human immunogenic
11	138	27.3	437	21	Human secreted pro

C	12	61	12.1	301	21	AAA06520	Human immunogenic
C	13	52.4	10.4	359	20	AAZ33445	Human prostate can
C	14	37.6	7.4	1515	22	AAF59232	Human PTH2 recepto
C	15	35.8	7.1	936	22	AAF58252	Oligonucleotide D1
C	16	35.8	7.1	936	22	AAF58254	Oligonucleotide D1
C	17	35.8	7.1	936	22	AAF58257	Oligonucleotide D1
C	18	35.8	7.1	936	22	AAF58259	Oligonucleotide D2
C	19	35.8	7.1	936	22	AAF58262	Oligonucleotide D2
C	20	35.8	7.1	938	22	AAF58255	Oligonucleotide D1
C	21	35.4	7.0	122186	22	AAZ09560	Human histone deac
C	22	35	6.9	753	18	AAZ84068	DNA encoding a S.
C	23	35	6.9	936	22	AAF58252	Oligonucleotide D1
C	24	35	6.9	936	22	AAF58254	Oligonucleotide D1
C	25	35	6.9	936	22	AAF58257	Oligonucleotide D1
C	26	35	6.9	936	22	AAF58259	Oligonucleotide D2
C	27	35	6.9	936	22	AAF58262	Oligonucleotide D2
C	28	35	6.9	938	22	AAF58255	Oligonucleotide D1
C	29	35	6.9	10207	15	AAO5134	Staphylococcus aur
C	30	34.8	6.9	982	20	AAZ37417	Human secreted pro
C	31	34.6	6.8	133719	21	AAZ64754	Macaca mulatta rha
C	32	34.4	6.8	119211	22	AAZ8553	Genomic fragment #
C	33	33.4	6.6	4594	18	AAV74511	Staphylococcus aur
C	34	33.2	6.6	3264	21	AAZ43195	Arabidopsis thalia
C	35	33.2	6.6	3333	21	AAZ49503	Arabidopsis thalia
C	36	33.2	6.6	11901	20	AAZ02998	Human IL-1ra BAC C
C	37	32.8	6.5	365	21	AAZ22045	Human secreted pro
C	38	32.6	6.4	46899	15	AAZ54386	T. niveum Cyclospo
C	39	32.4	6.4	578	21	AAA45319	Human secreted exp
C	40	32.4	6.4	6608	18	AAZ42751	Chicken CHD-1A gen
C	41	32.2	6.4	277	21	AAA00950	Human colon cancer
C	42	32.2	6.4	2427	20	AAZ99570	Nucleic acid sequ
C	43	31.6	6.3	847	21	AAZ34356	Arabidopsis thalia
C	44	31.6	6.2	348	20	AAZ51622	Human secreted pro
C	45	31.6	6.2	693	20	AAZ97741	Extended human sec

## ALIGNMENTS

RESULT 1	
ID	AAA06690/C
AA	AAA06690 standard; cDNA; 812 BP.
XX	
AC	AAA06690;
XX	
DT	13-JUN-2000 (first entry)
XX	
DE	Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:471.
XX	
KW	Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW	immunogenic; cytostatic; vaccine; ss.
XX	
OS	Homo sapiens.
XX	
PN	NO200004149-A2.
XX	
PD	27-JAN-2000.
XX	
PF	14-JUL-1999; 99MO-US15838.
XX	
PR	14-JUL-1998; 98US-0115453.
PR	14-JUL-1998; 98US-0116134.
PR	23-SEP-1998; 98US-0159812.
PR	23-SEP-1998; 98US-0159822.
PR	15-JAN-1999; 99US-0232149.
PR	15-JAN-1999; 99US-0232880.
PR	09-APR-1999; 99US-0288946.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Dillon DC, Harlocker SL, Yung J, Xu J, Mitcham JL;
XX	WPI, 2000-171268/15.

XX	New polypeptide useful for treating and diagnosing prostate cancer
PT	comprises an immunogenic portion of prostate tumor protein -
XX	
PS	Claim 1; Page 262; 263pp; English.
XX	
CC	The present invention describes isolated polypeptides, comprising an
CC	immunogenic portion of a prostate tumour protein (PP). The polypeptides
CC	and polynucleotides encoding them have cytostatic activity and can be
CC	used in vaccines and in gene therapy. The polypeptides and
CC	polynucleotides encoding them, antigen presenting cells which express
CC	the polypeptides, antibodies against the polypeptides and vaccines
CC	comprising them can be used for inhibiting the development of prostate
CC	cancer in a patient. The polypeptides can be used to generate antibodies
CC	or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC	the polynucleotides encoding the polypeptides can be used as a probe or
CC	to modulate the expression of the polypeptides. AA06741 to AA06891 and
CC	AA082000 to AA082020 represent sequences used in the exemplification of
CC	the present invention.
XX	
S0	Sequence 812 BP; 198 A; 189 C; 169 G; 256 T; 0 other;
	Query Match            48.1%; Score 243.4; DB 21; Length 812;
	Best Local Similarity   99.2%; Pred. No. 7.9e-70;
	Matches 255; Conservative 0; Mismatches 1; Indels 1; Gaps 1.
OY	250 cagacatcccttgaggagaatgcccgcgcccatcttgggtcattgatgagtcccgccctg 309
Db	541 CAGACATCCTCGTGGAATAATGCCGCCCCCATCTTGGGTCAATGATGAGCTCGCCCTG 482
OY	310 tgcctgttcctccgcttgtgaaggaaagacaattagaanaatlgaatgbtgtlccctaagg 369
Db	481 TGCCGTGTCGCCCTGTGTGAGGAGAAGACACTTAGAAAATGAATTGATGTCTCTTAAAG 422
OY	370 atggcagaagaaaacagatacctgttgtgatattttgaagcgatccaattgaaa 429
Db	421 ATGGCAGAGAAAACAGATCCCTGTGTGGATATTTATTTGAAGGGGTTTACAGATTTGANA 362
OY	430 tgaatccaccaagtgaatcaccaatgaagagaacacagacagaaaaatccttgatggc 489
Db	361 TGAAGTCA-CANAAGTGAGATTAATCAATGAGAGAAAAACAGACGAAAAATCTTGATGCG 303
OY	490 ttcaacaagacatgcaac 506
Db	302 TTCAACAAGACATGCAAC 286
RESULT 2	
AAV62429	
ID AAV62429 standard; CDNA: 820 BP.	
XX AAV62429;	
XX AC	
XX XX	
XX 30-DEC-1998 (first entry)	
XX DE	
DE Prostate cancer antigen (PCA3) cDNA splice variant 3.	
KW Prostate cancer antigen cDNA splice variant 3; PCA3; prostatic cancer;	
RW PC; ds.	
XX XX	
OS Homo sapiens.	
XX XX	
PN MO9845420-A1.	
PD 15-OCT-1998.	
XX XX	
PF 09-APR-1998; 98MO-CA00346.	
FR 10-APR-1997; 97US-0041836.	
XX XX	
PA (DIAG-) DIAGNOCURE INC.	

P1 Bussemakers MJG;  
xx  
DR WPI; 1998-568347/48.  
xx  
PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,  
PT prevention and treatment of prostatic cancer  
xx  
PS Claim 4; Pages 77-78; 111pp; English.

xx  
The present sequence represents the prostate cancer antigen (PCa3)  
CC cDNA splice variant 3 sequence comprising of exons 1, 3, and 4a  
CC of the PCa3 gene. The PCa3 cDNA splice variant 3 sequence,  
CC isolated from a human primary prostatic tumour tissue cDNA library,  
CC was found in approximately 15% of the cDNA clones isolated. The  
CC invention claims for PCa3 cDNA variants and the proteins they encode.  
CC The invention also claims for antibodies against PCa3 protein. The  
CC antibodies are claimed to be useful for detecting PCa3 protein in  
CC immunoassay tests, for diagnosing, assessing and prognosing of  
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin  
CC or radiolotope, and nucleic acids antisense to PCa3 cDNA are claimed  
CC to be useful for treating PC. While determining elevated levels of  
CC PCa3 (as RNA or protein) is useful for detecting a predisposition  
CC to development of PC, e.g. in prenatal tests. Detecting PCa3 protein  
CC allows differentiation between malignant and benign prostatic disease,  
CC and the level of PCa3 expression allows correlation with the grade of  
CC tumour. PCa3 protein and its fragments are also claimed to be useful  
CC in vaccines for preventing PC; in drug screens for identifying  
CC specific (ant)agonists (potentially useful therapeutically) and for  
CC studying protein-DNA interactions.

XX Sequence 820 BP; 262 A; 169 C; 191 G; 198 T; 0 other;

Query Match	48.1%	Score 243.4	DB 19	Length 820
Best Local Similarity	99.2%	Pred. No. 7.9e-70		
Matches 255	Conservative	0	Mismatches 1	Indels 1
OY	250	cacagatcccttggagaaatgcccgccgacatcttgggtcatcgatagcctgcctg	309	
Db	277	cagagatccctgggggaatgcccgcgcgacatcttgggtcatcgatagcctgcctg	336	
OY	310	tgccgtgctccgctgtgtgagggagagacattgaaaatgtaattgattgttccctaaagg	369	
Db	337	tgccgtgctccgctgtgtgagggagagacattgaaaatgtaattgattgttccctaaagg	396	
OY	370	atgggcagggaaaacagatccctgtgtgtgatattttatggaaaggagatttgaaa	429	
Db	397	atgggcagggaaaacagatccctgtgtgtgatattttatggaaaggagatttgaaa	456	
OY	430	tgaatgcccaaaagtgagcatcacaatggagggagaaaacagacggagaaacttgatggc	489	
Db	457	tgaatgcccaaaagtgagcatcacaatggagggagaaaacagacggagaaacttgatggc	515	
OY	490	ttcacagagatgtcaac	506	
Db	516	ttcacagagatgtcaac	532	
RESULT 3				
AAV62428				
ID	AAV62428 standard; cDNA; 1872 BP.			
XX	AAV62428;			
XX	30-DEC-1998 (first entry)			
XX	Prostate cancer antigen (PCA3) cDNA splice variant 2.			
XX	Prostate cancer antigen cDNA splice variant 2; PCA3; prostatic cancer;			
XX	PC; ds.			
XX	Homo sapiens.			

PM W09845420-A1.  
XX 15-OCT-1998.  
PD 09-APR-1998; 98WO-CA00346.  
XX 10-APR-1997; 97US-0041836.  
XX (DIAG-) DIAGNOCURE INC.  
XX Bussemakers MDG;  
XX WPI, 1998-568347/48.  
XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,  
PT prevention and treatment of prostatic cancer  
XX  
XX Claim 4; Pages 76-77; 11pp; English.  
XX  
XX The present sequence represents the prostate cancer antigen (PCA3)  
CC cDNA splice variant 2 sequence comprising of exons 1, 3, 4a and  
CC 4b of the PCA3 gene. The PCA3 cDNA splice variant 2 sequence,  
CC isolated from a human primary prostatic tumor tissue cDNA library,  
CC was found in approximately 65% of the cDNA clones isolated. The  
CC invention claims for PCA3 cDNA variants and the proteins they encode.  
CC The invention also claims for antibodies against PCA3 protein. The  
CC antibodies are claimed to be useful for detecting PCA3 protein in  
CC immunosay tests, for diagnosing, assessing and prognosing of  
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin  
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed  
CC to be useful for treating PC, while determining elevated levels of  
CC PCA3 (as RNA or protein) is useful for detecting a predisposition  
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein  
CC allows differentiation between malignant and benign prostatic disease,  
CC and the level of PCA3 expression allows correlation with the grade of  
CC tumor. PCA3 protein and its fragments are also claimed to be useful  
CC in vaccines for preventing PC; in drug screens for identifying  
CC specific (antagonists (potentially useful therapeutically) and for  
CC studying protein-DNA interactions.  
XX  
XX Sequence 1872 BP; 567 A; 389 C; 369 G; 539 T; 8 other;  
SQ

Query Match 48.1%; Score 243.4; DB 19; Length 1872;  
Best Local Similarity 99.2%; Pred. No. 1.2e-69;  
Matches 255; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 250 cacagatcccttggaagaatgcgcgcgcacatcttggtcatcatgagctgcgcctg 309  
DB 277 cagagatcccttggaagaatgcgcgcgcacatcttggtcatcatgagctgcgcctg 336  
QY 310 tgcctggtccgcttgtaggaagacattagaataatgatgtgttccttaag 369  
DB 337 tgcctggtccgcttgtaggaagacattagaataatgatgtgttccttaag 396  
QY 370 atggcagagaagaacagatctgtgtgatatatttgaaggagattcaatttga 429  
DB 397 atggcagagaagaacagatctgtgtgatatatttgaaggagattcaatttga 456  
QY 430 tgaagtcaacaagtgtgattacattagaaggaagaacagcaggaataattgtg 489  
DB 457 tgaagtca-caaagtgtgattacattagaaggaagaacagcaggaataattgtg 515  
QY 490 ttcaacaagatgcac 506  
DB 516 ttcaacaagatgcac 532

RESULT 4  
AAV62427 standard; cDNA; 2037 BP.  
XX  
XX AAV62427;

XX 30-DEC-1998 (first entry)  
DT Prostate cancer antigen (PCA3) cDNA splice variant 1.  
XX  
DE Prostate cancer antigen cDNA splice variant 1; PCA3; prostatic cancer;  
XX PC; ds.  
KM  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH 379..534  
FT GDS /\*tag= a  
FT /\*product= "PCA3 protein variant 1"  
FT 2019..2024  
FT polyA-signal /\*tag= b  
XX  
XX W09845420-A1.  
XX 15-OCT-1998.  
XX 09-APR-1998; 98WO-CA00346.  
XX 10-APR-1997; 97US-0041836.  
XX (DIAG-) DIAGNOCURE INC.  
XX Bussemakers MDG;  
XX WPI, 1998-568347/48.  
XX P-PSDB; AAV79736.  
XX  
XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,  
PT prevention and treatment of prostatic cancer  
XX  
XX Claim 3; Fig 2B-2J; 11pp; English.  
XX  
XX The present sequence represents the prostate cancer antigen (PCA3)  
CC cDNA splice variant 1 sequence comprising of exons 1, 2, 3, 4a and  
CC 4b of the PCA3 gene. The PCA3 cDNA splice variant 1 sequence,  
CC isolated from a human primary prostatic tumor tissue cDNA library,  
CC was found in approximately 5% of the cDNA clones isolated. The  
CC invention claims for PCA3 cDNA variants and the proteins they encode.  
CC The invention also claims for antibodies against PCA3 protein. The  
CC antibodies are claimed to be useful for detecting PCA3 protein in  
CC immunosay tests, for diagnosing, assessing and prognosing of  
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin  
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed  
CC to be useful for treating PC, while determining elevated levels of  
CC PCA3 (as RNA or protein) is useful for detecting a predisposition  
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein  
CC allows differentiation between malignant and benign prostatic disease,  
CC and the level of PCA3 expression allows correlation with the grade of  
CC tumor. PCA3 protein and its fragments are also claimed to be useful  
CC in vaccines for preventing PC; in drug screens for identifying  
CC specific (antagonists (potentially useful therapeutically) and for  
CC studying protein-DNA interactions.  
XX  
XX Sequence 2037 BP; 622 A; 426 C; 406 G; 575 T; 8 other;  
SQ

Query Match 48.1%; Score 243.4; DB 19; Length 2037;  
Best Local Similarity 99.2%; Pred. No. 1.3e-69;  
Matches 255; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 250 cacagatcccttggaagaatgcgcgcgcacatcttggtcatcatgagctgcgcctg 309  
DB 442 cagagatcccttggaagaatgcgcgcgcacatcttggtcatcatgagctgcgcctg 501  
QY 310 tgcctggtccgcttgtaggaagacattagaataatgatgtgttccttaag 369  
DB 502 tgcctggtccgcttgtaggaagacattagaataatgatgtgttccttaag 561

QY	370	atgggcgaggaataacagatccgctgtgtgataatatttcgaaggagatctaacgatttga	429
Db	562	atgggcgaggaataacagatccgctgtgtgataatatttcgaaggagatctaacgatttga	621
QY	430	tgaagtcaccacaatlgagcattaccaaatgagaggaaaacagacagaataatctgattgc	489
Db	622	tgaagtcac-caaagtgagcattaccaatgagaggaaaacagacagaataatctgattgc	680
QY	490	ttcacaagacatgcgaac	506
Db	681	ttcacaagacatgcgaac	697
RESULT	5		
AAA06688/c			
AAA06688	standard; cDNA; 2229 BP.		
AAA06688;			
13-JUN-2000	(first entry)		
Human	immunogenic prostate tumour protein cDNA sequence SEQ ID NO:469.		
Human:	prostate cancer; diagnosis; tumour; gene therapy; detection;		
immunogenic;	cytostatic; vaccine; ss.		
Homo sapiens.			
MO200004149-A2.			
27-JAN-2000.			
14-JUL-1999;	99MO-US15838.		
14-JUL-1998;	98US-0115453.		
14-JUL-1998;	98US-0116134.		
23-SEP-1998;	98US-0159812.		
23-SEP-1998;	98US-0159822.		
13-JAN-1999;	99US-0232149.		
15-JAN-1999;	99US-0232880.		
09-APR-1999;	99US-0288946.		
(CORI-) CORIXA CORP.			
Dillon DC, Harlocker SL, Yudin J, Xu J, Mitcham JL;			
WPI: 2000-171268/15.			
New polypeptide useful for treating and diagnosing prostate cancer			
comprises an immunogenic portion of prostate tumor protein -			
Claim 1; Page 260-261; 263pp; English.			
The present invention describes isolated polypeptides, comprising an			
immunogenic portion of a prostate tumour protein (PRP). The polypeptides			
and polynucleotides encoding them have cytostatic activity and can be			
used in vaccines and in gene therapy. The polypeptides and			
polynucleotides encoding them, antigen presenting cells which express			
the polypeptides, antibodies against the polypeptides and vaccines			
comprising them can be used for inhibiting the development of prostate			
cancer in a patient. The polypeptides can be used to generate antibodies			
or anti-idiotypic antibodies for passive immuno therapy. A portion of			
the polynucleotides encoding the polypeptides can be used as a probe or			
to modulate the expression of the polypeptides. AAA06241 to AAA06691 and			
AAV82000 to AAV82020 represent sequences used in the exemplification of			
the present invention.			
Sequence 2229 BP; 654 A; 447 C; 481 G; 647 T; 0 other;			

Query Match	48.1%	Score 243.4	DB 21	Length 2229
Best Local Similarity	99.28%	Pred. No. 1.4e-69		
Matches 255	Conservative	0	Mismatches 1	Indels 1
				Gaps 1

QY	250	cacagatcctctgggaggaatagcccgccgcacattgggtcaatgatatgacctgcctg	303
Db	1596	CAGAGATCCCTCGGGAGAAATGCCCGCGCCACCTTGGTGATCATGTAGAGCCCTGCCCTG	1537
QY	310	tgacctgtcccgctctgtgaggaagagacattagaataatgatgatgtcttccttaag	369
Db	1536	TGCCGCGCCGCTTGTGAGGGAAGGACATTAGAAATGATGATGTCTTCTTAAAG	1477
QY	370	atgggcgagaaaagatcctgttttggatattatttaacgggattacagattgaa	429
Db	1476	ATGGCGAGAAACAGATCCCTGTGTGATATTATTATTAACGGGATTCACATTTGAAA	1417
QY	430	tgaagtcaccaaagttagacattacaaatgagagaaacagacgaaatctltatggc	489
Db	1416	TGAAGTCA - CAAAGTGAGCATTCACATGAGAGAAACAGACGAAATCTTGATGGC	1358
QY	490	ttcacagaacatgtcaac	506
Db	1357	TTTCAAGACATGCAAC	1341

CC	RESULT 6
CC	AAA06689/c
XX	ID AAA06689 standard; cDNA; 2426 BP.
XX	
AC	AAA06689;
XX	
DT	13-JUN-2000 (first entry)
XX	
DE	Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:470.
XX	
KW	Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
XX	immunogenic; cytosolic; vaccine; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200004149-A2.
XX	
PD	27-JAN-2000.
XX	
PF	14-JUL-1999; 99WO-US15838.
XX	
PR	14-JUL-1998; 98US-0115453.
XX	
PR	14-JUL-1998; 98US-0116134.
XX	
PR	23-SEP-1998; 98US-0159812.
XX	
PR	23-SEP-1998; 98US-0159822.
XX	
PR	15-JAN-1999; 99US-0232149.
XX	
PR	15-JAN-1999; 99US-0232880.
XX	
PR	09-APR-1999; 99US-0288946.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;
XX	
DR	WPI; 2000-171268/15.
XX	
PT	New polypeptide useful for treating and diagnosing prostate cancer
XX	comprises an immunogenic portion of prostate tumor protein -
PS	Claim 1; Page 261-262; 263pp; English.
XX	
XX	The present invention describes isolated polypeptides, comprising an
CC	immunogenic portion of a prostate tumour protein (PMP). The polypeptides
CC	and polynucleotides encoding them have cytosolic activity and can be
CC	used in vaccines and in gene therapy. The polypeptides and
CC	polynucleotides encoding them, antigen presenting cells which express
CC	the polypeptides, antibodies against the polypeptides and vaccines
CC	comprising them can be used for inhibiting the development of prostate
CC	cancer in a patient. The polypeptides can be used to generate antibodies
CC	or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC	the polynucleotides encoding the polypeptides can be used as a probe or
CC	to modulate the expression of the polypeptides. AAA06241 to AAA06691 and



CC AAY82000 to AAY82020 represent sequences used in the exemplification of  
CC the present invention.

SO Sequence 2426 BP; 717 A; 476 C; 548 G; 685 T; 0 other;

Query Match 48.1%; Score 243.4; DB 21; Length 2426;  
Best Local Similarity 99.2%; Pred. No. 1.4e-69;  
Matches 255; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 250 cacaagatcccttgaggaatgcccgcgcgcacatcttggtgcatcgatgagcctcgccctg 309  
DB 1590 CAGACATCCCTGCGAATAATGCCGCCCATCTTGCTGATCATGAGCCTCGCCCTG 1531  
QY 310 tgcctggtccgccttggtgaggaaggaacattagaaatgaaatgattgtgtccctaaag 369  
DB 1530 TGCCGTGTCCTCCCTGTGTGAGGAGACATAGAAATGAAATGATGTCTTAAAG 1471  
QY 370 atggcaggaagaaacagatctgtgtgatatatttgaaggatatacaattgaa 429  
DB 1470 ATGGGACAGAAACAGATCTCTGTGATATTATTTGACGGATTTACAGATTTGAAA 1411  
QY 430 tgaagtcaccaagtgcagcatcaccatgagagaaacagacgagaaatcttgatgc 489  
DB 1410 TGAAGTCA-CAAAAGTGAGCATTTACCATGAGAGAAACAGACGAGAAATCTTGATGCGC 1352  
QY 490 ttcaagaacatgcacac 506  
DB 1351 TTCACAGACATGCAAC 1335

## RESULT 7

AAA06687  
ID AAA06687 standard; cDNA; 3112 BP.

AC AAA06687;

DT 13-JUN-2000 (first entry)

DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:468.

KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;

KW Immunogenic; cytostatic; vaccine; ss.

OS Homo sapiens.

PN WO200004149-A2.

PD 27-JAN-2000.

PF 14-JUL-1999; 99WO-US15838.

PR 14-JUL-1998; 98US-0115453.

PR 14-JUL-1998; 98US-0116134.

PR 23-SEP-1998; 98US-0159812.

PR 15-JAN-1999; 98US-0159822.

PR 15-JAN-1999; 98US-0232149.

PR 15-JAN-1999; 98US-0232880.

PR 09-APR-1999; 99US-0288946.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Harlocker SL, Yudin J, Xu J, Mitcham JL;

DR WPI: 2000-171268/15.

XX New polypeptide useful for treating and diagnosing prostate cancer  
XX comprises an immunogenic portion of prostate tumor protein -  
XX Claim 1; Page 259-260; 263pp; English.  
XX The present invention describes isolated polypeptides, comprising an  
XX immunogenic portion of a prostate tumour protein (PTP). The polypeptides

CC and polynucleotides encoding them have cytostatic activity and can be  
CC used in vaccines and in gene therapy. The polypeptides and  
CC polynucleotides encoding them, antigen presenting cells which express  
CC the polypeptides, antibodies against the polypeptides and vaccines  
CC comprising them can be used for inhibiting the development of prostate  
CC cancer in a patient. The polypeptides can be used to generate antibodies  
CC or anti-idiotype antibodies for passive immuno therapy. A portion of  
CC the polynucleotides encoding the polypeptides can be used as a probe or  
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and  
CC AAY82000 to AAY82020 represent sequences used in the exemplification of  
CC the present invention.

SO Sequence 3112 BP; 975 A; 587 C; 624 G; 926 T; 0 other;

Query Match 48.1%; Score 243.4; DB 21; Length 3112;  
Best Local Similarity 99.2%; Pred. No. 1.6e-69;  
Matches 255; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 250 cacaagatcccttgaggaatgcccgcgcgcacatcttggtgcatcgatgagcctcgccctg 309  
DB 1492 CAGACATCCCTGCGAATAATGCCGCCCATCTTGCTGATCATGAGCCTCGCCCTG 1551  
QY 310 tgcctggtccgccttggtgaggaaggaacattagaaatgaaatgattgtgtccctaaag 369  
DB 1552 TGCCGTGTCCTCCCTGTGTGAGGAGACATAGAAATGAAATGATGTCTTAAAG 1611  
QY 370 atggcaggaagaaacagatctgtgtgatatatttgaaggatatacaattgaa 429  
DB 1612 ATGGGACAGAAACAGATCTCTGTGATATTATTTGACGGATTTACAGATTTGAAA 1671  
QY 430 tgaagtcaccaagtgcagcatcaccatgagagaaacagacgagaaatcttgatgc 489  
DB 1672 TGAAGTCA-CAAAAGTGAGCATTTACCATGAGAGAAACAGACGAGAAATCTTGATGCGC 1730  
QY 490 ttcaagaacatgcacac 506  
DB 1731 TTCACAGACATGCAAC 1747

## RESULT 8

AAV62430  
ID AAV62430 standard; cDNA; 3582 BP.

AC AAV62430;

DT 30-DEC-1998 (first entry)

DE Prostate cancer antigen (PCA3) wild-type cDNA.

KW Prostate cancer antigen cDNA; PCA3; prostatic cancer;

KW PC; ds.

OS Homo sapiens.

PN CDS

FT Key Location/Qualifiers

FT CDS 401..556

FT /tag= a

FT /product= "PCA3 protein"

FT polyA\_signal 983..987

FT polyA\_signal 2041..2046

FT polyA\_signal 2597..2602

FT polyA\_signal 3494..3496

FT /tag= e

XX WO9845420-A1.

XX 15-OCT-1998.

XX 09-APR-1998; 98WO-CA00346.

```
XX 10-APR-1997; 97US-0041836.
XX (DIAG-) DIAGNOCURE INC.
XX Bussemakers MUG;
XX MPI; 1998-568347/48.
XX P-PSDB; AAW79738.
XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
XX prevention and treatment of prostatic cancer
XX
XX Claim 3; Fig 5B-5F; 11pp; English.
XX
XX The present sequence represents the prostate cancer antigen (PCA3)
XX wild-type cDNA sequence comprising of exons 1, 2, 3, 4a-4d of the
XX PCA3 gene. The invention claims for PCA3 cDNA variants and the
XX proteins they encode. The invention also claims for antibodies
XX against PCA3 protein. The antibodies are claimed to be useful for
XX detecting PCA3 protein in immunoassay tests, for diagnosing, assessing
XX and prognosing of prostatic cancer (PC). Antibodies, optionally
XX coupled to a cytotoxin or radioisotope, and nucleic acids antisense
XX to PCA3 cDNA are claimed to be useful for treating PC, while determining
XX elevated levels of PCA3 (as RNA or protein) is useful for detecting a
XX predisposition to development of PC, e.g. in prenatal tests. Detecting
XX PCA3 protein allows differentiation between malignant and benign
XX prostatic disease, and the level of PCA3 expression allows correlation
XX with the grade of tumour. PCA3 protein and its fragments are also
XX claimed to be useful in vaccines for preventing PC; in drug screens
XX for identifying specific (ant)agonists (potentially useful
XX therapeutically) and for studying protein-DNA interactions.
XX
XX Sequence 3582 BP; 1052 A; 788 C; 679 G; 1063 T; 0 other;
XX
Query Match 48.1%; Score 243.4; DB 19; Length 3582;
Best Local Similarity 99.2%; Pred. No. 1,7e-69;
Matches 255; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
OY 250 cacgagatcccgaggagaatgccgcgcgcatctgtggtcatgagctgcgacctg 309
DB 464 cagagatcccgaggagaatgccgcgcgcatctgtggtcatgagctgcgacctg 523
OY 310 tgcctgtcccgctgtgtaggaagagacattagaagaatgatgtgttcctaaagg 369
DB 524 tgcctgtcccgctgtgtaggaagagacattagaagaatgatgtgttcctaaagg 583
OY 370 atgggcaggaagaacagatcctgtgtgtgattatttgaaacgggattacagattgaaa 429
DB 584 atgggcaggaagaacagatcctgtgtgtgattatttgaaacgggattacagattgaaa 643
OY 430 tgaagtcacaaagtgtgacattacaaagtgaagaacacagacgagaacatttgatgc 489
DB 644 tgaagtcacaaagtgtgacattacaaagtgaagaacacagacgagaacatttgatgc 702
OY 490 ttcaacaagacatgcaac 506
DB 703 ttcaacaagacatgcaac 719
XX
RESULT 9
AAK37486
ID AAK37486 standard; cDNA; 597 BP.
XX
XX AAK37486;
XX 06-JUL-1999 (first entry)
XX Human secreted protein cDNA fragment containing gene 36.
XX
XX Human secreted protein; treatment; prevention; protein therapy; AIDS;
XX gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
XX
```

```
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW arthritis; malignancy; digestive; endocrine; infection; ss.
XX
XX Homo sapiens.
XX
XX MO9918208-A1.
XX
XX 15-APR-1999.
XX
XX 01-OCT-1998; 98MO-US20775.
XX
XX 02-OCT-1997; 97US-0060884.
XX 02-OCT-1997; 97US-0060833.
XX 02-OCT-1997; 97US-0060836.
XX 02-OCT-1997; 97US-0060837.
XX 02-OCT-1997; 97US-0060838.
XX 02-OCT-1997; 97US-0060839.
XX 02-OCT-1997; 97US-0060843.
XX 02-OCT-1997; 97US-0060862.
XX 02-OCT-1997; 97US-0060866.
XX 02-OCT-1997; 97US-0060874.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Carter KC, Duan DR, Endress GA, Feng P, Ferlie AM;
XX Florence KA, Greene JM, Janat F, Lafleur DW, Ni J;
XX Rosen CA, Ruben SM, Shi Y, Young P, Yu G;
XX
XX MPI; 1999-264022/22.
XX P-PSDB; AAY07887.
XX
XX New isolated human genes and the secreted polypeptides they encode
XX
XX Claim 1a; Page 247; 368pp; English.
XX
XX This invention describes novel isolated human genes and the secreted
XX proteins they encode. The products of the invention are useful for
XX preventing, treating or ameliorating medical conditions, e.g. by protein
XX or gene therapy. Also pathological conditions can be diagnosed by
XX determining the amount of the new polypeptides in a sample or by
XX determining the presence of mutations in the new polynucleotides.
XX Specific uses are described for each of the 101 polynucleotides, based
XX on which tissues they are most highly expressed in, and include
XX developing products for the diagnosis or treatment of cancer, tumours,
XX neurodegenerative disorders, developmental abnormalities and fetal
XX deficiencies, blood disorders, leukemias, diseases of the immune system,
XX autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
XX allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
XX disease, skeletal or cardiac muscle disorders, pulmonary disorders,
XX transplant rejection, disorders involving osteoclasts such as
XX osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
XX infections and AIDS. The human secreted proteins of the invention are
XX represented in AAY07887-107993 and the encoding nucleic acids are
XX represented in AAK37451-K37552.
XX
XX Sequence 597 BP; 181 A; 131 C; 134 G; 150 T; 1 other;
XX
Query Match 44.4%; Score 224.6; DB 20; Length 597;
Best Local Similarity 99.2%; Pred. No. 9,9e-64;
Matches 235; Conservative 1; Mismatches 0; Indels 1; Gaps 1;
OY 270 gccggcgccattgttgatgacatgagctgcgcctgtgctgtgctgcgcttgag 329
DB 54 gccggcgccattgttgatgacatgagctgcgcctgtgctgtgctgcgcttgag 113
OY 330 ggaagacattagaagaatgatgtgttccttaagaagatggcgagaaacagatcc 389
DB 114 ggaagacattagaagaatgatgtgttccttaagaagatggcgagaaacagatcc 173
XX
```

Qy	390	tggtgtgatttatttgaaacgggattacgatttgaaatgaagtcacaaagtgagca	449
Dy	174	tgctgtgatttatttgaaacgggwtctacgatttgaaatgaagtcacaaagtgagca	232
Qy	450	ttaccatgaagaggaacacagacgagaaatctgatgcttcacaaacatgacac	506
Dy	233	ttaccatgagaggaacacagacgagaaatctgatgcttcacaaacatgacac	289
Db			
RESULT	10		
AC	AAA06545	standard; cDNA; 718 BP.	
XX	AAA06545;		
XX	13-JUN-2000	(first entry)	
XX			
DE	Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:313.		
XX			
KW	Human; prostate cancer; diagnosis; tumour; gene therapy; detection;		
XX	immunogenic; cytosolic; vaccine; ss.		
OS	Homo sapiens.		
XX			
XX	MO200004149-A2.		
PD			
XX	27-JAN-2000.		
XX			
PF	14-JUL-1999; 99WO-US15838.		
XX			
PR	14-JUL-1998; 98US-0115453.		
PR	14-JUL-1998; 98US-0116134.		
PR	23-SEP-1998; 98US-0159812.		
PR	23-SEP-1998; 98US-0159822.		
PR	15-JAN-1999; 99US-0232149.		
PR	15-JAN-1999; 99US-0232880.		
PR	09-APR-1999; 99US-0288946.		
XX			
PA	(CORI-) CORIXA CORP.		
XX			
PI	Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;		
XX			
DR	WPI; 2000-171268/15.		
XX			
PT	New polypeptide useful for treating and diagnosing prostate cancer		
PT	comprises an immunogenic portion of prostate tumor protein -		
XX			
PS	Claim 1; Page 199-200; 263pp; English.		
XX			
CC	The present invention describes isolated polypeptides, comprising an		
CC	immunogenic portion of a prostate tumour protein (PTP). The polypeptides		
CC	and polynucleotides encoding them have cytostatic activity and can be		
CC	used in vaccines and in gene therapy. The polypeptides and		
CC	polynucleotides encoding them, antigen presenting cells which express		
CC	the polypeptides, antibodies against the polypeptides and vaccines		
CC	comprising them can be used for inhibiting the development of prostate		
CC	cancer in a patient. The polypeptides can be used to generate antibodies		
CC	or anti-idiotypic antibodies for passive immuno therapy. A portion of		
CC	the polynucleotides encoding the polypeptides can be used as a probe or		
CC	to modulate the expression of the polypeptides. AAA06241 to AAA06691 and		
CC	AAV82000 to AAV82020 represent sequences used in the exemplification of		
CC	the present invention.		
XX			
XX			
Sequence	718 BP; 222 A; 145 C; 169 G; 172 T; 10 other;		
Query Match	43.2%; Score 218.4; DB 21; Length 718;		
Best Local Similarity	98.1%; Pred. No. 1.2e-61;		
Matches	252; Conservative 0; Mismatches 2; Indels 3; Gaps 3;		
Yy	250	cacagatccctggagaaatgccgcgcacatcttggtcatgatgatgagctgcctg	309

Dd	252	caggaatccctcggaggaaabgccccgccgcacattcttggtcatcgatgaaccttcgccctt	311
Qy	310	tgcctcgtcccctctgtgagggaaagacatagaanaatgatgatgtgtctcttaag	369
Dd	312	tgcctgntcccgcttgtgaggaagacattagaanaatgatgatgtgtctcttaag	371
Qy	370	atggcgcggaaaacaagatccctgtttgttgatatattttagaacgggatataagtgttgaa	429
Dd	372	at-ggcagaaaaacaagaccctgtgtgtgatatattttagaacgggatatacaagttaa	430
Qy	430	tgaagtcaccaagtgtgacattaccattgaaaggaacacagacagaaactctgatgac	489
Dd	431	tgaagtcac-caaatgtgacattaccattgaaaggaacacagacagaaactctgatg-	488
Qy	490	ttcacaaagacatgacac	506
Dd	489	ttcacaaagacatgacac	505
<hr/>			
RESULT 11			
AAAC06768	ID	AAAC06768 standard; cDNA; 437 BP.	
XX	AC	AAAC06768;	
XX	DT	06-OCT-2000 (first entry)	
XX	DE	Human secreted protein 5' EST, SEQ ID NO: 10843.	
KW	HM	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;	
OS	XX	gene therapy; chromosome mapping; ss.	
PN	EP	Homo sapiens.	
XX	PM	EP1033401-A2.	
PD	PF	06-SEP-2000.	
XX	Pf	21-FEB-2000; 2000EP-0200610.	
PR	XX	26-FEB-1999; 99US-0122487.	
PA	(JEST )	GENSET.	
PI	Dumas Milne Edwards J,	Duclet A, Giordano J;	
DR	WP:	2000-500381/45.	
XX	PT	New nucleic acid that is a 5' expressed sequence tag (5' EST) for	
PT	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for		
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures -		
PS	Claim 1;	SEQ ID 10843; 71pp + CD-ROM; English.	
CC	The present sequence is one of a large number of 5' ESTs derived from		
CC	mRNAs encoding secreted proteins. No ORF has yet been conclusively		
CC	identified within the present sequence. The 5' ESTs were prepared from		
CC	total human RNAs or polyA+ RNAs derived from 30 different tissues. EST		
CC	sequences usually correspond mainly to the 3' untranslated region (UTR)		
CC	of the mRNA because they are often obtained from oligo-dT primed cDNA		
CC	libraries. Such ESTs are not well suited for isolating cDNA sequences		
CC	derived from the 5' ends of mRNAs and even in those cases where longer		
CC	cDNA sequences have been obtained, the full 5' UTR is rarely included.		
CC	5' ESTs are derived from mRNAs with intact 5' ends and can therefore be		
CC	used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used		
CC	in diagnostic, forensic, gene therapy and chromosome mapping procedures.		
CC	They are used to obtain upstream regulatory sequences and to design		
CC	expression and secretion vectors.		
XX	Sequence 437 BP; 140 A; 87 C; 118 G; 88 T; 4 other;		
S0			

Query Match      27.3%; Score 138; DB 21; Length 437;

Best Local Similarity 97.9%; Pred. No. 2; 2e-35; Matches 18; Conservative 1; Mismatches 2; Indels 0; Gaps 0

OY 250 cacaagatcccttggaagaataatgcccgcgccacatcctgggtcatcatgatgactgcacctg 309  
|||  
Db 297 cagaagatcccttggagaaatgatalccccgcgcaccattctgggtcatcatgatgactgcacctg 356  
|||  
OY 310 tgcctgttcctcgctttggaaggagaacatttagaaaatgaatgatgttgttccttaag 369  
|||||  
Db 357 tgcctgttcctcgctttggaaggagaacatttagaaaatgaatgatgttgttccttaag 416  
|||||  
OY 370 atgggcagaaaaacagatcct 390  
|||||  
Db 417 atgggcagaaaaacagatcct 437  
|||||

RESULT 12  
AAA06520/c  
ID AAA06520 standard; cDNA; 301 BP.  
XX  
AC AAA06520;  
XX  
DT 13-JUN-2000 (first entry)  
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:287.  
XX  
KM Human; prostate cancer; diagnosis; tumour; gene therapy; detection;  
KW Immunogenic; cytosolic; vaccine; ss.  
OS Homo sapiens.  
XX  
XX WO200004149-A2.  
PN 27-JAN-2000.  
PD 14-JUL-1999; 99WO-US15838.  
XX  
PE 14-JUL-1998; 98US-0115453.  
PR 14-JUL-1998; 98US-0116134.  
PR 23-SEP-1998; 98US-0159812.  
PR 23-SEP-1998; 98US-0159822.  
PR 15-JAN-1999; 99US-0232149.  
PR 15-JAN-1999; 99US-0232880.  
PR 09-APR-1999; 99US-0288946.  
XX  
PA (CORI-) CORIXA CORP.  
PI Dillion DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;  
XX WPI: 2000-171268/15.  
DR  
XX  
XX  
PT New polypeptide useful for treating and diagnosing prostate cancer  
XX comprises an immunogenic portion of prostate tumor protein -  
XX  
PS Claim 1; Page 192; 263pp; English.

The present invention describes isolated polypeptides, comprising an immunogenic portion of a prostate tumor protein (TRP). The polypeptides and polynucleotides encoding them have cytostatic activity and can be used in vaccines and in gene therapy. The polypeptides and polynucleotides encoding them, antigen presenting cells which express the polypeptides, antibodies against the polypeptides and vaccines comprising them can be used for inhibiting the development of prostate cancer in a patient. The polypeptides can be used to generate antibodies or anti-idiotypic antibodies for passive immuno therapy. A portion of the polynucleotides encoding the polypeptides can be used as a probe or to modulate the expression of the polypeptides. AAA06241 to AAA06591 and AA062000 to AA062020 represent sequences used in the exemplification of the present invention.

Sequence 301 BP; 76 A; 58 C; 70 G; 97 T; 0 other;

	Query Match	12.1%	Score 61:	DB 21:	Length 301:	
	Best Local Similarity	100.0%	Pred. No. 3.4e-10:			
	Matches 61:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:	
QY	446 agcctaccatgtgagagaacacagacagagaaactctgatgcgttcacaagacatgcaa 505       301 AGCATTACCATGTGAGAGAAACAACAGACAGAATACTTGATGGCTTCCACAGACATGCCAA 242					
QY	506 c 506 					
Db	241 C 241					
RESULT 13						
ID	AAZ33445					
AC	AAZ33445 standard; CDNA: 359 BP.					
XX						
DT	08-DEC-1999 (first entry)					
XX						
DE	Human prostate cancer-associated EST 23.					
XX						
KW	Expressed sequence tag; EST; prostate tumor; antitumor; treatment; gene therapy; tissue specificity human; ss.					
OS	Homo sapiens.					
XX						
PN	DE1981193-AI.					
PD	16-SEP-1999.					
XP						
PP	10-MAR-1998; 98DE-1011193.					
XX						
PR	10-MAR-1998; 98DE-1011193.					
XX						
PA	(META-) METAGEN GES GENOMFORSCHUNG MBH.					
PI						
PI	Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;					
XX						
DR	WPI: 1999-519628/44.					
DR	P-PDSB: AAAY48243.					
PT	New nucleic acid expressed at high level in prostatic tumor tissue and					
PT	encoded polypeptides, useful for treating cancer and screening for					
PS	therapeutic agents .					
PS	Claim 1a; 87; 16pp; German.					
XX						
CC	This invention describes novel nucleic acid sequences (A) that are					
CC	expressed at high level in prostatic tumor tissue and encode gene					
CC	products or their fragments. The products of the invention have					
CC	antitumor activity. Polypeptides (I) encoded by (A) are used: (1) for					
CC	identifying agents for treatment of prostatic cancer and (II) for					
CC	therapy of prostate cancer, optionally where expressed by gene therapy					
CC	methods. (A) is also used to isolate full-length genes (for gene therapy)					
CC	and for recombinant production of (I), which can be used to raise					
CC	specific antibodies. (A) are identified by assembly of ESTs (expressed					
CC	sequence tags) before they are analyzed for expression pattern (tissue					
CC	specificity). This approach eliminates many of the false results, as					
CC	regards tissue specificity, associated with known methods that use					
CC	single (usually short) ESTs. AAZ33423-23476 represent expressed					
CC	sequence tags described in the method of the invention.					
XX						
SQ	Sequence 359 BP; 121 A; 75 C; 94 G; 69 T; 0 other;					
Query Match	10.4%;	Score 52.4;	DB 20;	Length 359;		
Best Local Similarity	98.1%;	Pred. No. 2.5e-07;				
Matches 53:	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;		
Y	250 cacagatccctggaggaaatgcccgcccaacctcttggtcatcatgatagacctc 303 					

Db 306 cagcgcacccctggcggaagaatgcccgcgcgcattgttgcatgatgagctc 359

RESULT 14  
ID AAF59232/c  
AAF59232 standard; DNA; 1515 BP.

AC AAF59232;  
XX  
XX 26-APR-2001 (first entry)  
DE Human PTH2 receptor ligand TIP39 related genomic DNA sequence.

KW Tuberin/fundibular peptide of 39 residues; TIP39; PTH2 receptor ligand;  
KM parathyroid hormone; parathyroid hormone receptor; anorectic; nootropic;  
KM PTH2 receptor binding activity; antidepressant; neuroleptic; analgesic;  
KM antimigraine; antidiabetic; osteopathic; hypertensive; cardiatic;  
KM cytosstatic; antiaslathmatic; neuroprotective; PTH receptor antagonist;  
KM obesity; eating disorder; metabolic disorder; mental disorder;  
KM depression; schizoprenia; dementia; acute pain; chronic pain; migraine;  
KM headache; diabetes; osteoporosis; hypercalcaemia; calcium metabolism;  
KM hyperension; congestive heart failure; tumour; asthma; emphysema;  
KM restrictive lung disease; demyelinating condition; multiple sclerosis;  
KW leukodystrophy; ds.

OS Homo sapiens.  
XX  
XX WO200077042-A2.  
XX PD  
XX 21-DEC-2000.  
XX PF 15-JUN-2000; 2000WO-US16776.  
XX XX 15-JUN-1999; 99US-0139335.  
PR (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PA  
XX Usdin TB, Hoare SRJ;  
PI  
XX WPI: 2001-122833/13.  
DR  
XX New parathyroid hormone type 2 or 1 receptor ligand, useful for  
PT treating e.g. migraine or headaches, hypertension, obesity and other  
PT eating or metabolic disorders, mental disorders and osteoporosis -  
XX  
XX Example 9; Page 78-79; 106pp; English.

CC The present invention describes an isolated or purified peptide (I) that  
CC is a parathyroid hormone type 2 (PTH2) receptor ligand (Ia) or PTH1  
CC receptor ligand (Ib). (I) has anorectic, neuroleptic, antidepressant,  
CC nootropic, analgesic, antimigraine, antidiabetic, osteopathic,  
CC hypertensive, cardiatic, cytosstatic, antiaslathmatic and neuroprotective  
CC activities, and is an PTH receptor antagonist. The peptide is useful in  
CC the treatment of a disorder mediated by binding a PTH2 or PTH1 receptor.  
CC The PTH2 receptor binding activity may be used in treating obesity or  
CC other eating or metabolic disorders, mental disorders (e.g. depression,  
CC schizophrenia and dementia), acute or chronic pain, migraine or  
CC headaches, diabetes and other metabolic disorders, osteoporosis,  
CC hypercalcaemia and other disorders affecting calcium metabolism,  
CC hypertension, congestive heart failure and control of tumour growth,  
CC asthma, emphysema or other restrictive lung diseases, and demyelinating  
CC conditions such as multiple sclerosis and leukodystrophies. The present  
CC sequence represents a PTH2 receptor ligand TIP39 related genomic DNA  
CC sequence, which is given in the exemplification of the present invention.  
XX  
XX Sequence 1515 BP; 269 A; 451 C; 518 G; 277 T; 0 other;

SQ Query Match 7.4%; Score 37.6; DB 22; Length 1515;  
Best Local Similarity 50.0%; Pred. No. 0.039;  
Matches 94; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

83 tctctactcgtttctactccttactactactgltcttcgcgaatacctaaccagtatttc 142

[illegible]



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 16, 2001, 17:57:37 ; Search time 51.36 Seconds  
(without alignments)  
27.149 Million cell updates/sec

Title: US-09-675-650-3  
Perfect score: 123  
Sequence: 1 MFHISPEKYPHTQKRAQR 23

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.GeneSeq 0601.\*  
1: /SIDSB/gcgdata/geneSeq/geneSeq/AA1980.DAT.\*  
2: /SIDSB/gcgdata/geneSeq/geneSeq/AA1981.DAT.\*  
3: /SIDSB/gcgdata/geneSeq/geneSeq/AA1982.DAT.\*  
4: /SIDSB/gcgdata/geneSeq/geneSeq/AA1983.DAT.\*  
5: /SIDSB/gcgdata/geneSeq/geneSeq/AA1984.DAT.\*  
6: /SIDSB/gcgdata/geneSeq/geneSeq/AA1985.DAT.\*  
7: /SIDSB/gcgdata/geneSeq/geneSeq/AA1986.DAT.\*  
8: /SIDSB/gcgdata/geneSeq/geneSeq/AA1987.DAT.\*  
9: /SIDSB/gcgdata/geneSeq/geneSeq/AA1988.DAT.\*  
10: /SIDSB/gcgdata/geneSeq/geneSeq/AA1989.DAT.\*  
11: /SIDSB/gcgdata/geneSeq/geneSeq/AA1990.DAT.\*  
12: /SIDSB/gcgdata/geneSeq/geneSeq/AA1991.DAT.\*  
13: /SIDSB/gcgdata/geneSeq/geneSeq/AA1992.DAT.\*  
14: /SIDSB/gcgdata/geneSeq/geneSeq/AA1993.DAT.\*  
15: /SIDSB/gcgdata/geneSeq/geneSeq/AA1994.DAT.\*  
16: /SIDSB/gcgdata/geneSeq/geneSeq/AA1995.DAT.\*  
17: /SIDSB/gcgdata/geneSeq/geneSeq/AA1996.DAT.\*  
18: /SIDSB/gcgdata/geneSeq/geneSeq/AA1997.DAT.\*  
19: /SIDSB/gcgdata/geneSeq/geneSeq/AA1998.DAT.\*  
20: /SIDSB/gcgdata/geneSeq/geneSeq/AA1999.DAT.\*  
21: /SIDSB/gcgdata/geneSeq/geneSeq/AA2000.DAT.\*  
22: /SIDSB/gcgdata/geneSeq/geneSeq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	123	100.0	51	AAW79736	Prostate cancer an
2	123	100.0	51	AAW79738	Prostate cancer an
3	56	45.5	11	AAW79737	Prostate cancer an
4	46.5	37.8	119	AAW96038	80 kD protein. Te
5	44.5	36.2	234	AAW20288	Vesicular stomatit
6	44.5	36.2	419	AAW08021	A micropthalma a
7	44.5	36.2	511	AAW04539	Vesiculovirus glyc
8	44.5	36.2	511	AAW23748	Vesicular stomatit
9	44.5	36.2	511	AAW73506	VSV-G protein sequ
10	44.5	36.2	511	AAW20290	Vesicular stomatit
11	44.5	36.2	511	AAW59298	Vesicular stomatit

12	44	35.8	492	19	AAW56281	Babesia microti BM
13	44	35.8	492	20	AAW24342	Babesia microti an
14	44	35.8	492	21	AAW30191	B. microti BMN1-4
15	44	35.8	503	19	AAW56302	Babesia microti MN
16	44	35.8	503	20	AAW24364	Babesia microti an
17	44	35.8	503	21	AAW30216	B. microti clone a
18	43.5	35.4	166	21	AAW5637	Human colon cancer
19	43.5	35.4	245	8	AAW70677	Human kallikrein g
20	43.5	35.4	245	10	AAW92314	Human recombinant k
21	43.5	35.4	256	21	AAW21315	Human KKL1. Homo
22	43.5	35.4	262	8	AAW70568	Human kallikrein-1
23	43.5	35.4	262	10	AAW95121	Kallikrein encoded
24	43.5	35.4	262	19	AAW71005	Human prostate-ass
25	43.5	35.4	262	21	AAW21319	Human KKL2. Homo
26	43.5	35.4	267	21	AAW54293	Human pancreatic c
27	43.5	35.4	511	22	AAW20289	Vesicular stomatit
28	43	35.0	177	21	AAW85637	Monkey fibroblast
29	43	35.0	178	21	AAW85638	Human fibroblast g
30	43	35.0	178	21	AAW85639	Human fibroblast g
31	43	35.0	250	21	AAW43790	Arabidopsis thalia
32	43	35.0	251	22	AAW65297	Human PRO9828 Prot
33	43	35.0	363	21	AAW77125	Human neurotransmi
34	43	35.0	585	20	AAW23643	Protein encoded by
35	42	34.1	127	20	AAW24088	Salmonella typhimu
36	42	34.1	188	21	AAW58237	Arabidopsis thalia
37	42	34.1	232	21	AAW58236	Arabidopsis thalia
38	42	34.1	247	21	AAW58235	Arabidopsis thalia
39	41.5	33.7	364	18	AAW36443	Thermotabile alpha
40	41.5	33.7	431	12	AAW13731	Human cyclin A. H
41	41.5	33.7	515	21	AAW56707	Human prostate can
42	41	33.3	119	11	AAW09304	Sequence deduced f
43	41	33.3	167	21	AAW22828	Arabidopsis thalia
44	41	33.3	178	21	AAW41814	Human ORF1578
45	41	33.3	250	21	AAW94965	Human secreted pro

#### ALIGNMENTS

RESULT 1	
ID	AAW79736 standard; Protein; 51 AA.
AAW79736:	
AC	30-DEC-1998 (first entry)
XX	
DT	Prostate cancer antigen (PCA3) protein variant 1.
XX	
DE	Prostate cancer antigen protein variant 1; PCA3; prostatic cancer;
KW	PC.
KW	
OS	Homo sapiens.
PN	W09845420-A1.
PD	15-OCT-1998.
PF	09-APR-1998; 98WO-CA00346.
PR	10-APR-1997; 97US-0041836.
PA	(DIAG-) DIAGNOCURE INC.
PI	Bussemakers MGJ;
PI	WPI: 1998-568347/48.
DR	N-PSDB; AAW62427.
XX	
PT	New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
PS	prevention and treatment of prostatic cancer
PS	Claim 16; Fig 2B-2J; 111pp; English.

45 108 #1 cultured

XX The present sequence represents the prostate cancer antigen (PCA3)  
CC protein variant 1 encoded by a PCA3 spliced cDNA sequence comprising  
CC of exons 1, 2, 3, 4a and 4b of the PCA3 gene. The invention claims  
CC for PCA3 cDNA variants and the proteins they encode. The invention  
CC also claims for antibodies against PCA3 protein. The antibodies are  
CC claimed to be useful for detecting PCA3 protein in immunosay tests,  
CC for diagnosing, assessing and prognosing of prostatic cancer (PC).  
CC Antibodies, optionally coupled to a cytotoxin or radioisotope, and  
CC nucleic acids antisense to PCA3 cDNA are claimed to be useful for  
CC treating PC, while determining elevated levels of PCA3 (as RNA or  
CC protein) is useful for detecting a predisposition to development of  
CC PC, e.g. in prenatal tests. Detecting PCA3 protein allows  
CC differentiation between malignant and benign prostatic disease,  
CC and the level of PCA3 expression allows correlation with the grade of  
CC tumour. PCA3 protein and its fragments are also claimed to be useful  
CC in vaccines for preventing PC; in drug screens for identifying  
CC specific (ant)agonists (potentially useful therapeutically) and for  
CC studying protein-DNA interactions.

XX Sequence 51 AA:

Query Match 100.0%; Score 123; DB 19; Length 51;  
Best Local Similarity 100.0%; Pred. No. 9.7e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MELHISPFKYPHTQEAQKEAQR 23  
|  
Db 1 mflhispfkyphtqeqakeqr 23

RESULT 2

AAW79738  
ID AAW79738 standard; Protein; 51 AA.

AC AAW79738;

DT 30-DEC-1998 (first entry)

DE Prostate cancer antigen (PCA3) wild-type protein.

XX Prostate cancer antigen protein variant 1; PCA3; prostatic cancer;  
KW PC.

XX Homo sapiens.

OS MO9845420-A1.

PN 15-OCT-1998.

PD 09-APR-1998; 98WO-CA00346.

PF 10-APR-1997; 97US-0041836.

PR (DIAG-) DIAGNOCURE INC.

PA Bussemakers MJG;

PI WPI: 1998-568347/48.

DR N-PSDB; AAV62430.

XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,  
PT prevention and treatment of prostatic cancer

PS Claim 16; Fig 5B-5F; 11pp; English.

XX The present sequence represents the prostate cancer antigen (PCA3)  
CC wild-type protein sequence encoded by a PCA3 wild-type cDNA sequence  
CC comprising of exons 1, 2, 3, 4a-4d of the PCA3 gene. The invention  
CC claims for PCA3 cDNA variants and the proteins they encode. The  
CC invention also claims for antibodies against PCA3 protein. The  
CC antibodies are claimed to be useful for detecting PCA3 protein in

CC immunosay tests, for diagnosing, assessing and prognosing of  
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin  
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed  
CC to be useful for treating PC, while determining elevated levels of  
CC PCA3 (as RNA or protein) is useful for detecting a predisposition to  
CC development of PC, e.g. in prenatal tests. Detecting PCA3 protein  
CC allows differentiation between malignant and benign prostatic disease,  
CC and the level of PCA3 expression allows correlation with the grade of  
CC tumour. PCA3 protein and its fragments are also claimed to be useful  
CC in vaccines for preventing PC; in drug screens for identifying  
CC specific (ant)agonists (potentially useful therapeutically) and for  
CC studying protein-DNA interactions.

XX Sequence 51 AA:

Query Match 100.0%; Score 123; DB 19; Length 51;  
Best Local Similarity 100.0%; Pred. No. 9.7e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MELHISPFKYPHTQEAQKEAQR 23  
|  
Db 1 mflhispfkyphtqeqakeqr 23

RESULT 3  
AAW79737  
ID AAW79737 standard; Peptide; 11 AA.

AC AAW79737;

DT 30-DEC-1998 (first entry)

DE Prostate cancer antigen (PCA3) protein antigenic peptide.

XX Prostate cancer antigen cDNA splice variant 1; PCA3; prostatic cancer;  
KW PC; PCA3 antigenic peptide.

XX Homo sapiens.

OS MO9845420-A1.

PN 15-OCT-1998.

PD 09-APR-1998; 98WO-CA00346.

PF 10-APR-1997; 97US-0041836.

PR (DIAG-) DIAGNOCURE INC.

PA Bussemakers MJG;

PI WPI: 1998-568347/48.

DR New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,  
PT prevention and treatment of prostatic cancer

PS Disclosure; Page 21; 11pp; English.

XX The present sequence represents an antigenic peptide fragment  
CC derived from prostate cancer antigen (PCA3) protein. The invention  
CC claims for PCA3 cDNA variants and the proteins they encode. The  
CC invention also claims for antibodies against PCA3 protein. The  
CC antibodies are claimed to be useful for detecting PCA3 protein in  
CC immunosay tests, for diagnosing, assessing and prognosing of  
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin  
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed  
CC to be useful for treating PC, while determining elevated levels of  
CC PCA3 (as RNA or protein) is useful for detecting a predisposition  
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein  
CC allows differentiation between malignant and benign prostatic disease,  
CC and the level of PCA3 expression allows correlation with the grade of  
CC tumour. PCA3 protein and its fragments are also claimed to be useful





CC colon carcinoma, hematopoietic tumour, glioblastoma, melanoma,  
 CC sarcoma, fibrosarcoma, neuroendocrine tumour, acute myelogenous  
 CC leukaemia, chronic myelogenous leukaemia, promyelocytic leukaemia,  
 CC lymphoma and melanoma (all claimed).

SO Sequence 234 AA;

Query Match 36.2%; Score 44.5; DB 22; Length 234;  
 Best Local Similarity 42.9%; Pred. No. 20;  
 Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

OY 3 LHISSP---FKYPHTQEAQKE 20  
 11:11 1:11 1:11  
 Db 135 lhlsskagvfphidgaag 155

# RESULT 6

AAB08021 standard; Protein; 419 AA.

AC AAB08021;

DT 14-NOV-2000 (first entry)

DE A microphthalmia-associated transcription factor (Mtf).

KW Microphthalmia-associated transcription factor; Mtf; Mtf(+); Mtf(-);  
 splice variant; cancer cell; cancer cell marker; melanoma.

XX Homo sapiens.

XX Key Location/Qualifiers

FT MISC-difference 187..192  
 FT "these amino acids are uniquely present in  
 splice variant Mtf(+)"

WT WO200047765-A1.

PN 17-AUG-2000.

PE 03-FEB-2000; 2000WO-GB00313.

PR 08-FEB-1999; 99GB-0002776.

PA (CURI-) CURIE CANCER CARE MARIE.

PI Coding CR;

XX WPI: 2000-549153/50.

DR N-PSDB; AAA59714.

XX Novel method for the detection of melanoma comprises determining the  
 PT predominance of the splice variants Mtf(-) and Mtf(+) in a  
 PT biological sample -

PS Disclosure; Fig 1; 29pp; English.

XX The present sequence represents a microphthalmia-associated transcription  
 CC factor (Mtf). The protein exists in 2 forms, Mtf(+) and Mtf(-).

CC Amino acids 187-192 are uniquely present in the Mtf(+) splice variant.  
 CC In normal cells, Mtf(+) RNA is predominantly present, while in certain  
 CC cancer cell lines, Mtf(-) is predominantly present. The specification  
 CC describes a method for determining the predominance of either of the  
 CC splice variants Mtf(-) and Mtf(+) in a biological sample. The method  
 CC comprises independently detecting Mtf(-) and Mtf(+), and then  
 CC comparing the amounts of the spliced variants in order to determine  
 CC which variant is predominant in the sample. The method is used to detect  
 CC the presence of and determine the predominance of the two splice variants  
 CC of the cancer cell marker Mtf. Determination of these parameters leads  
 CC to an indication of the presence, or likelihood of developing cancer,  
 CC especially a melanoma.

SO Sequence 419 AA;

Query Match 36.2%; Score 44.5; DB 21; Length 419;  
 Best Local Similarity 47.1%; Pred. No. 39;  
 Matches 8; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

OY 4 HISSPFKYPHTQEAQKE 20  
 1:11 1:11 1:11  
 Db 15 hlenptky-hlqgaqrq 30

# RESULT 7

AAW04539 standard; Protein; 511 AA.

AC AAW04539;

DT 26-JAN-1997 (first entry)

DE Vesiculovirus glycoprotein (G).

KW Rhabdovirus; VSV; vaccine; pVSVFL(+); vector; immunotherapy;  
 antigen; tumour; vesiculovirus.

XX Vesicular stomatitis virus.

XX Key Location/Qualifiers

FT Domain 456..482  
 FT /label= "Transmembrane\_domain  
 483..511  
 FT /label= Cytoplasmic\_domain

PN WO9634625-A1.

PD 07-NOV-1996.

PE 01-MAY-1996; 96WO-US06053.

PR 04-MAY-1995; 95US-0435032.

PA (UYVA ) UNIV YALE.

PI Rose JK;

XX WPI: 1996-505901/50.

DR N-PSDB; AAT38744.

XX Recombinant replicable vesiculovirus - useful to develop prods. for  
 PT diagnosis and therapy, partic. in vaccines for infections, diseases  
 PT or tumours

PS Disclosure; Page 91-92; 162pp; English.

XX A single glycoprotein (G) species (AAW04540) spans the membrane of  
 CC vesiculovirus (VSV) and forms spikes on the surface of the virus  
 CC particle. 3 Proteins, termed N or nucleocapsid (AAW04536), P  
 CC (formerly termed NS indicating non-structural) (AAW04537) and L or  
 CC large (AAW04540) are associated with the VSV nucleocapsid. An  
 CC additional matrix (M) protein (AAW04538) lies within the membrane  
 CC envelope. The proteins are encoded by VSV(-) DNA in pVSVFL(+)  
 CC (AAT38744), a plasmid that can be used in the prodn. of recombinant,  
 CC replicable VSV for use as vaccines.

XX Sequence 511 AA;

SO Sequence 511 AA;

Query Match 36.2%; Score 44.5; DB 17; Length 511;  
 Best Local Similarity 42.9%; Pred. No. 48;  
 Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

OY 3 LHISSP---FKYPHTQEAQKE 20  
 11:11 1:11 1:11

Db 412 lhlsskayvfehphidqdaasg 432

# RESULT 8

AAV23748 standard; Protein; 511 AA.

AAV23748;

08-SEP-1999 (first entry)

Vesicular stomatitis virus G stem polypeptide.

Rhabdovirus; heterologous fusion protein; virus fusion; target cell membrane; N protein; P protein; L protein; G stem polypeptide; antireceptor protein; infection; human immune deficiency virus; neoplastic condition; benign tumour; polyp; cafe au lait spot; leukoplakia; skin mole.

Vesicular stomatitis virus.

WO9932648-A1.

01-JUL-1999.

22-DEC-1998; 98WO-US26084.

22-DEC-1997; 97US-0068472.

(UYTE-) UNIV TENNESSEE RES CORP.

Robinson CS, Whitt MA;

WPI; 1999-418938/35.

Recombinant rhabdovirus containing heterologous fusion protein

Disclosure; Page 14; 63pp; English.

The specification describes genetically engineered rhabdoviruses which comprise a heterologous F (fusion) protein, or its fragment, to facilitate fusion of the virus with a target cell membrane, and the rhabdovirus N, P and L proteins. Recombinant viruses which include a G stem polypeptide and an antireceptor protein (ARP) are used to fuse with target cells (depending on nature of ARP), for treatment of diseases associated with a surface protein recognized by ARP. Particularly the method is used to treat viral, parasitic and bacterial infections (specifically human immune deficiency virus); (pre-)neoplastic conditions; benign tumours; polyps; cafe au lait spots; leukoplakia and skin moles. Cells may be treated in vivo or in vitro. Recombinant rhabdoviruses that infect diseased or abnormal cells and also express a reporter can be used for diagnosis, detection and monitoring progression/regression of disease, also for studying function and specificity of proteins not naturally found on rhabdovirus. The present sequence may be used to create the rhabdoviruses of the invention.

Sequence 511 AA;

Query Match 36.2%; Score 44.5; DB 20; Length 511;

Best Local Similarity 42.9%; Pred. No. 48; Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

3 LHISSP---FKYPHTQEAQKE 20

412 lhlsskayvfehphidqdaasg 432

# RESULT 9

AAW73506 standard; Protein; 511 AA.

AAW73506;

01-MAR-1999 (first entry)

VSV-G protein sequence.

DAF; decay accelerating factor; human; complement protein; gene therapy;

viral vector; VSV-G protein.

Vesicular stomatitis virus.

JP10313865-A.

02-DEC-1998.

15-MAY-1997; 97JP-0125965.

15-MAY-1997; 97JP-0125965.

(DINA-) DINABEKKU KENKUYUSHO KK.

WPI; 1999-074147/07.

N-PSDB; AAV08936.

Vector having complement controlling factor - useful for gene therapy

Example 3; Page 9-11; 15pp; Japanese.

This sequence represents the VSV-G protein. DNA encoding this sequence can be used in the viral vector of the invention. The viral vector contains a factor controlling the function of a human complement protein, particularly a membrane combining type protein. The viral vector, which is stable in situ, is useful for gene therapy.

Sequence 511 AA;

Query Match 36.2%; Score 44.5; DB 20; Length 511;

Best Local Similarity 42.9%; Pred. No. 48; Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

3 LHISSP---FKYPHTQEAQKE 20

412 lhlsskayvfehphidqdaasg 432

# RESULT 10

AAAB20290 standard; Protein; 511 AA.

AAAB20290;

29-MAY-2001 (first entry)

Vesicular stomatitis virus mutant M4 G protein.

VSV; G protein; tumour; ovarian carcinoma; lung carcinoma; colon carcinoma; leukaemia; lymphoma; myeloma; glioblastoma; melanoma; sarcoma; neuroendocrine tumour; cancer; therapy; oncolytic; cytostatic; antitumour; mutant; mutein.

Vesicular stomatitis virus.

Key Location/Qualifiers

Misc-difference 294..348

/label= Unknown

/note= "encoded by incompletely sequenced portion of N gene"

Misc-difference 254

/note= "Glu in wild-type strain"

MO200119380-A2.



PR 01-OCT-1996; 96US-0723142.  
XX  
XX (CORI-) CORIXA CORP.  
PI Houghton R, Lodes MJ, Reed SG, Sleath PR;  
XX WPI: 1998-195465/18.  
DR N-PSDB; AAY22734.  
XX  
XX polypeptides comprising Babesia microti antigens and their  
PT immunogenic fragments or epitopes - and related nucleic acid,  
PT vectors, transformed cells and antibodies, useful for diagnosis of  
PT infection and in protective vaccines  
XX  
PS Claim 1; Page 49-51; 113pp; English.  
XX  
XX The sequence is that of a polypeptide comprising at least  
CC one antigenic portion of a Babesia microti antigen. It can be used  
CC to diagnose B. microti infection by detecting specific antibodies  
CC in usual immunoassays. Infection can also be diagnosed using:  
CC (a) primers or probes derived from the coding sequence, in  
CC standard amplification or hybridisation tests, or (b) using  
CC antibodies to detect the corresponding antigen. It is also  
CC useful in vaccines to protect against infection, especially  
CC when formulated with an adjuvant. The new diagnostic methods  
CC allow rapid differentiation between B. microti infection and  
CC other tick-borne diseases (Lyme disease and ehrlichiosis) that  
CC have similar symptoms but require different treatments.  
XX  
SQ Sequence 492 AA;

Query Match 35.8%; Score 44; DB 19; Length 492;  
Best Local Similarity 31.8%; Pred. No. 56;  
Matches 7; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 FLHISPFKYPHTEAKKEAQR 23  
| : | | : : | | : : | : : :  
Db 103 fryiskeyehtelakenck 124

RESULT 13  
AAY24342  
ID AAY24342 standard; Protein; 492 AA.  
XX  
XX AAY24342;  
AC  
XX  
XX 16-SEP-1999 (first entry)  
DT  
XX  
XX Babesia microti antigen BMN1-4.  
DE  
XX  
XX Babesia microti; antigen; immunogen; diagnosis; infection; vaccine;  
KW immunity; detection.  
XX  
XX Babesia microti.  
OS  
XX  
XX WO9929869-A1.  
PN  
XX  
XX 17-JUN-1999.  
PD  
XX  
XX 11-DEC-1998; 98MO-US26437.  
PF  
XX  
XX 11-DEC-1997; 97US-0990571.  
PR  
XX  
XX (CORI-) CORIXA CORP.  
PA (MAYO-) MAYO FOUNDATION.  
XX  
XX Bruinsma E, Houghton R, Lodes MJ, Persing D, Reed SG;  
PI Sleath PR;  
XX  
XX WPI: 1999-385612/32.  
DR N-PSDB; AAX88998.  
XX

PT New isolated Babesia microti polypeptides  
XX  
XX Example 1; Page 72-74; 126pp; English.  
PS  
XX  
XX The present invention describes isolated polypeptides comprising  
CC specific immunogenic portions of Babesia microti. AAX88983 to AAX88994  
CC encode specifically claimed B. microti immunogenic proteins, and  
CC AAY24327 to AAY24338 represent these proteins. B. microti polypeptides  
CC and nucleic acids can be used for detecting B. microti infections. They  
CC can also be used in vaccines for inducing protective immunity against B.  
CC microti infections. The present sequence represents a B. microti antigen.  
XX  
SQ Sequence 492 AA;

Query Match 35.8%; Score 44; DB 20; Length 492;  
Best Local Similarity 31.8%; Pred. No. 56;  
Matches 7; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 FLHISPFKYPHTEAKKEAQR 23  
| : | | : : | | : : | : : :  
Db 103 fryiskeyehtelakenck 124

RESULT 14  
AAB30191  
ID AAB30191 standard; Protein; 492 AA.  
XX  
XX AAB30191;  
AC  
XX  
XX 12-FEB-2001 (first entry)  
DT  
XX  
XX B. microti BMN1-4 antigen SEQ ID NO: 21.  
DE  
XX  
XX Babesiosis; rodent parasite; tick-borne illness; antigen;  
KW disease diagnosis; disease prevention.  
XX  
XX Babesia microti.  
OS  
XX  
XX WO200060090-A1.  
PN  
XX  
XX 12-OCT-2000.  
PD  
XX  
XX 05-APR-2000; 2000MO-US09136.  
PF  
XX  
XX 05-APR-1999; 99US-0286488.  
PR  
XX  
XX 17-MAR-2000; 2000US-0528784.  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX  
XX Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD;  
PI WPI: 2000-686939/67.  
DR N-PSDB; AAC65081.  
XX

PT New polypeptides containing an antigenic portion of Babesia microti  
PT antigen and DNAs encoding the polypeptides, useful for diagnosing,  
PT treating or preventing B. microti infection, or for inducing protective  
PT immunity in a patient  
XX  
XX Disclosure; Page 72-73; 118pp; English.  
PS  
XX  
XX The present invention is related to the isolation of antigenic sequences  
CC from the rodent parasite Babesia microti. This organism is transmitted to  
CC humans by the same tick which transmits Lyme disease and ehrlichiosis.  
CC The organism causes a malaria-like infection known as babesiosis. The  
CC sequences identified by this invention can be used in the diagnosis,  
CC prevention and treatment of babesiosis.  
XX  
XX  
SQ Sequence 492 AA;

Query Match 35.8%; Score 44; DB 21; Length 492;

Best Local Similarity 31.8%; Pred. NO. 56;  
Matches 7; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Job time: 3896 sec

OY 2 FLHISPPFKYPHTQEAQKEAQR 23  
| : | | : | | : | : : :  
Db 103 frylskeyeyentelakehckk 124

## RESULT 15

AAW56302  
ID AAW56302 standard; Protein; 503 AA.

XX AAW56302;

DT 28-SEP-1998 (first entry)

DE Babesia microti MN-10 antigen sequence.

XX antigen; detection; diagnosis; vaccine; tick-borne disease;

KW differentiation; Lyme disease; ehrlichiosis.

XX Babesia microti.

OS EP834567-A2.

PD 08-APR-1998.

PF 01-OCT-1997; 97EP-0117067.

PR 24-APR-1997; 97US-0845258.

PR 01-OCT-1996; 96US-0723142.

XX (CORI-) CORIXA CORP.

PI Houghton R, Lodes MJ, Reed SG, Sleath PR;

DR WPI; 1998-195465/18.

DR N-PSDB; AAV22748.

PT Polypeptides comprising Babesia microti antigens and their  
PT immunogenic fragments or epitopes - and related nucleic acid,  
PT vectors, transformed cells and antibodies, useful for diagnosis of  
PT infection and in protective vaccines

PS Claim 1; Page 99-101; 113pp; English.

CC The sequence is that of a polypeptide comprising at least  
CC one antigenic portion of a Babesia microti antigen. It can be used  
CC to diagnose B. microti infection by detecting specific antibodies  
CC in usual immunoassays. Infection can also be diagnosed using:  
CC (a) primers or probes derived from the coding sequence, in  
CC standard amplification or hybridisation tests, or (b) using  
CC antibodies to detect the corresponding antigen. It is also  
CC useful in vaccines to protect against infection, especially  
CC when formulated with an adjuvant. The new diagnostic methods  
CC allow rapid differentiation between B. microti infection and  
CC other tick-borne diseases (Lyme disease and ehrlichiosis) that  
CC have similar symptoms but require different treatments.

XX Sequence 503 AA:

Query Match 35.8%; Score 44; DB 19; Length 503;  
Best Local Similarity 31.8%; Pred. NO. 57;

Matches 7; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

OY 2 FLHISPPFKYPHTQEAQKEAQR 23  
| : | | : | | : | : : :  
Db 31 frylskeyeyentelakehckk 52



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2001, 17:51:07 ; Search time 1566.45 Seconds

(without alignments)  
4996.445 Million cell updates/sec

Title: US-09-675-650-1

Perfect score: 506  
Sequence: 1 caggaaagcacaaggaagc.....gacctacaagacatgcaac 506

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 segs, 773874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_cm:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pi1:\*  
13: gb\_pi2:\*  
14: gb\_pi3:\*  
15: gb\_pi4:\*  
16: em\_ba1:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rnd:\*  
22: em\_htg\_hum1:\*  
23: em\_htg\_hum2:\*  
24: em\_htg\_hum3:\*  
25: em\_htg\_hum4:\*  
26: em\_htg\_hum5:\*  
27: em\_htg\_hum6:\*  
28: em\_htg\_hum7:\*  
29: em\_htg\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_other:\*  
33: em\_htg\_rnd:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_cm:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pi:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_v1:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_v11:\*  
59: gb\_v12:\*  
60: gb\_hcg1:\*  
61: gb\_hcg2:\*  
62: gb\_hcg3:\*  
63: gb\_hcg4:\*  
64: gb\_hcg5:\*  
65: gb\_hcg6:\*  
66: gb\_hcg7:\*  
67: gb\_hcg8:\*  
68: gb\_hcg9:\*  
69: gb\_hcg10:\*  
70: gb\_hcg11:\*  
71: gb\_hcg12:\*  
72: gb\_hcg13:\*  
73: gb\_hcg14:\*  
74: gb\_hcg15:\*  
75: gb\_hcg16:\*  
76: gb\_hcg17:\*  
77: gb\_hcg18:\*  
78: gb\_hcg19:\*  
79: gb\_hcg20:\*  
80: gb\_hcg21:\*  
81: gb\_hcg22:\*  
82: gb\_hcg23:\*  
83: gb\_hcg24:\*  
84: gb\_hcg25:\*  
85: gb\_pr1:\*  
86: gb\_pr2:\*  
87: gb\_pr3:\*  
88: gb\_pr4:\*  
89: gb\_pr5:\*  
90: gb\_pr6:\*  
91: gb\_pr7:\*  
92: gb\_pr8:\*  
93: gb\_pr9:\*  
94: gb\_rot1:\*  
95: gb\_rot2:\*  
96: gb\_in4:\*  
97: gb\_pi10:\*  
98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	494	97.6	164371	80	AL390239	AL390239 Homo sapi
2	494	97.6	173831	80	AL359314	AL359314 Homo sapi
3	494	97.6	267581	80	AL358573	AL358573 Homo sapi
4	482	95.3	5435	88	AF103808	AF103808 Homo sapi
5	243.4	48.1	3923	88	AF103907	AF103907 Homo sapi
6	58.2	11.5	7218	10	166494	166494 Sequence 14
7	52.4	10.4	359	9	AX018075	AX018075 Sequence
8	43	8.5	102200	92	HS30M3	AT031775 Human DNA

```

C 9 42 8.3 215103 72 AC058821 Homo sapi
C 10 41.2 8.1 9070 13 AF148542 Arabidops
C 11 175391 69 AC025622 Mus muscu
C 12 40.8 8.1 132887 90 AL160408 Human DNA
C 13 40 7.9 199103 76 AC080076 Homo sapi
C 14 39.8 7.9 166774 85 AC005225 Homo sapi
C 15 39.4 7.8 127091 61 AC009843 Drosophila
C 16 39.4 7.8 156736 86 AC006502 Homo sapi
C 17 39.4 7.8 164331 87 AC016254 Homo sapi
C 18 39.4 7.8 166300 66 AC021373 Homo sapi
C 19 39.4 7.8 171419 86 AC006961 Homo sapi
C 20 39.4 7.8 191148 70 AC027588 Homo sapi
C 21 39.4 7.8 222895 76 AC084065 Mus muscu
C 22 39.2 7.7 160910 83 AP001929 Homo sapi
C 23 39.2 7.7 184800 75 AC074144 Mus muscu
C 24 39.2 7.7 209346 70 AC026373 Homo sapi
C 25 38.8 7.7 17609 5 AF208112 Trypanoso
C 26 38.8 7.7 118030 82 AP000667 Homo sapi
C 27 38.8 7.7 158759 63 AC013484 Trypanoso
C 28 38.8 7.7 164587 67 AC023001 Homo sapi
C 29 38.8 7.7 196361 85 AC005386 ciltb-57-1
C 30 38.8 7.7 203407 60 AC006174 Homo sapi
C 31 38.6 7.6 88782 67 AC022625 Homo sapi
C 32 38.4 7.6 159350 68 AC024332 Homo sapi
C 33 38.4 7.6 200125 71 AC034105 Homo sapi
C 34 38.2 7.5 150748 68 AC024096 Homo sapi
C 35 38.2 7.5 164485 69 AC026106 Homo sapi
C 36 38.2 7.5 171154 83 AP003356 Homo sapi
C 37 38.2 7.5 192465 83 AP002851 Homo sapi
C 38 38.0 7.5 153342 82 AL589949 Homo sapi
C 39 38.0 7.5 158373 73 AC068912 Mus muscu
C 40 38.0 7.5 236406 76 AC079489 Mus muscu
C 41 37.8 7.5 2292 8 DL4316 D14316 delta-cryst
C 42 37.8 7.5 19479 94 M2A21XCOA 222923 M.musculus
C 43 37.8 7.5 134210 85 AC005052 Homo sapi
C 44 37.8 7.5 159818 60 AC007337 Homo sapi
C 45 37.8 7.5 163229 87 AC009039 Homo sapi

```

## ALIGNMENTS

```

RESULT 1
AL390239 164371 bp DNA HTG 09-MAR-2001
LOCUS Homo sapiens chromosome 9 clone RP11-58J3, *** SEQUENCING IN
DEFINITION PROGRESS ***, 4 unordered pieces.
ACCESSION AL390239.11 GI:13274794
VERSION Homo sapiens HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS human.
SOURCE Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
ORGANISM Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE 1 (bases 1 to 164371)
AUTHORS Burton, J.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Mar 12, 2001 this sequence version replaced gi:13273805.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
Contact: humquerry@sanger.ac.uk
Center project name: BA58J3
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid: L08752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 16337 bases at least Q40

```

```

Consensus quality: 163481 bases at least Q30
Consensus quality: 163577 bases at least Q20
Insert size: 164071; sum-of-contigs
Insert size: 166918; 4.2% error; agarose-fp
Quality coverage: 8.78x in Q20 bases; sum-of-contigs Quality
coverage: 9.07x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 20241 20340: gap of 100 bp
* 20341 116100: contig of 95760 bp in length
* 116101 116200: gap of 100 bp
* 116201 118243: contig of 2043 bp in length
* 118244 118343: gap of 100 bp
* 118344 164371: contig of 46028 bp in length.
location/Qualifiers
1. 164371
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-58J3"
/clone_1ib="RPC1-11.1"
/clone_1ib="RPC1-11.1"
1. 20240
/note="assembly: fragment: 00838
fragment_chain: 1
clone_end: SP6
vector_side: left"
20341. 116100
/note="assembly: fragment: 01998
fragment_chain: 1"
116201. 118243
/note="assembly: fragment: 00916
fragment_chain: 1"
118344. 164371
/note="assembly: fragment: 00334
fragment_chain: 1
clone_end: T7
vector_side: right"

```

```

BASE COUNT 50254 a 34905 c 32663 g 46249 t 300 others
ORIGIN

```

```

Query Match 97.6%; Score 494; DB 80; Length 164371;
Best Local Similarity 99.8%; Pred. No. 1.2e-134;
Matches 505; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
OY 1 caggagacacaaaggaagcagagagtaagtgtctataaagaactcaatttactca 60
Db 33067 CAGGAGACACAAAGGAAGCAGAGGTAAGTCTTATAAGACACATTTCTACTCA 33126
OY 61 gaattttgtgtcctaaattctctaccgcgtttcatcctctactcaactgtctc 120
Db 33127 GAAATTTTGTATGGCCTTAAATCTTCTACGTTTCTATCTTCTACTGACGTGCTC 33186
OY 121 ccggaatccactaccgaattctctctccctcgtatgtctgaactgtcacttga 180
Db 33187 CCGGATCCACTACCGATTTTCTATTCTTTCGTCGTAATTTCTGACAGGCTCACTGGA 33246
OY 181 ttatcctcaagagtcgtgatttctacccgggtcaccctcgtccctcatttctgc 240
Db 33247 TTTATCCCTACAGGAGTCTGATTTTCTACCGGCTCACCTCCCTCCATATTTGTC 33306
OY 241 ctccacttccagatccctgtggagaatgcggcgccatcttggatcgatgagc 300
Db 33307 CTCACATTTCACAGATCCCTGTGGAGAAATGCCGCGCCCATCTTGGGTCAATGATGAC 33366

```

QY 301 ctcgcccctgtgcctgtgcccgcctgtgtgaagggaagacattagaanaatgatgatgtgt 360  
|||||  
Db 33367 CTCGCCCTGTGCTGCTGCTGCCGCTGTGTGAGGAGGAGCATTAAGAAATGATTCATCTGTT 33426  
QY 361 ccttaaggatgtggcaggaagaacagatcctgtgtgtgatatatttgaacgggattaca 420  
|||||  
Db 33427 CCTTAAAGATGGGCGAGGAAACAGATCCTGTGTGTGATTTATTTGAAACGGGATTACA 33486  
QY 421 gatttgaatgaagtacccaagttagcattaccattgagaggaacagacagaagaat 480  
|||||  
Db 33487 GATTGTAAATGAAAGTCA-CAAGGTGAGCATTTACCAATGAGAGGAAACAGACGAGAAAT 33545  
QY 481 ctgtatgcttcaacaagacatgcaac 506  
|||||  
Db 33546 CTTGATGGCTTCACAAGACATGCAAC 33571  
RESULT 2  
AL359314 173831 bp DNA HTG 08-APR-2001  
LOCUS Homo sapiens chromosome 9 clone RP11-1081L4, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, 3 unordered pieces.  
ACCESSION AL359314  
VERSION AL359314.12 GI:13396560  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_ACTIVERFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 173831)  
Direct Submission  
Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Mar 20, 2001 this sequence version replaced gi:13277120.  
----- Genome Center  
Center: Sanger Centre  
Web site: http://www.sanger.ac.uk  
Contact: humquerry@sanger.ac.uk  
----- Project Information  
Center project name: BA1081L4  
----- Summary Statistics  
Sequencing program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 173545 bases at least Q40  
Consensus quality: 173599 bases at least Q30  
Consensus quality: 173621 bases at least Q20  
Insert size: 173631; sum-of-contrigs  
Insert size: 172123; 10.0% error; agarose-fp  
Quality coverage: 10.35x in Q20 bases; sum-of-contrigs Quality  
coverage: 10.50x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contrigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contrigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 50595: contig of 50595 bp in length  
\* 50596 50695: gap of 100 bp  
\* 50696 157636: contig of 106941 bp in length  
\* 157637 157736: gap of 100 bp  
\* 157737 173831: contig of 16095 bp in length.  
Location/Qualifiers  
1. 173831  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="9"

/clone="RP11-1081L4"  
/clone.lib="RPC1-11.1"  
1. 50595  
/note="assembly-fragment:04137  
fragment\_chain:1"  
50696. 157636  
/note="assembly-fragment:04236  
fragment\_chain:1"  
157737. 173831  
/note="assembly-fragment:01556"  
BASE COUNT 52107 a 36469 c 35018 g 50037 t 200 others  
ORIGIN  
Query Match 97.6%; Score 494; DB 80; Length 173831;  
Best Local Similarity 99.8%; Pred. No. 1.2e-134;  
Matches 505; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 1 caggagacacaaaggagacagagtaagtcttataaagcactcaattctactca 60  
|||||  
Db 155988 CAGGAGACACAAAGAGACACAGAGTAAGTCTTTATRAAGCACTCAATTTCTACTCA 156047  
QY 61 gaaattttgatgtgccttaagtccctactactcgtttctatcccttactactcactc 120  
|||||  
Db 156048 GAAATTTTGTGATGGCTTAAAGTTCCTTAAGTCTGTTCTATACCTTCTACTGACACTGTCCTC 156107  
QY 121 ccggaatccactacagatttctatttcttgcctgtatgtctgtacgtgctacttgga 180  
|||||  
Db 156108 CCGGAATCCACATCCATTTTATTTCTTCTGCTCATATGTGTGACTGCTCATTGGA 156167  
QY 181 ttatctcaccagagctcgtgatttctaccggcgcacccctccctccatattgttc 240  
|||||  
Db 156168 TTTATCTTCAAGAGATCTGTGATTTTCTACCCGGGCTCACCCTCCCTCCATATTTGTC 156227  
QY 241 ctcaccattcacagatccctgtgagagaatgcccgcgcacatcttggatcagatgagc 300  
|||||  
Db 156228 CTCACCTTTCACAGATCCCTGGAGAAATGCCGCCGCATCTTGCGTCATGATGAGAC 156287  
QY 301 ctcgcccctgtgcctgtgcccgcctgtgtgaagggaagacattagaanaatgatgatgtgt 360  
|||||  
Db 156288 CTCGCCCTGTGCTGCTGCTGCCGCTGTGTGAGGAGGAGCATTAGAAATGATTCATCTGTT 156347  
QY 361 ccttaaggatgtggcaggaagaacagatcctgtgtgtgatatatttgaacgggattaca 420  
|||||  
Db 156348 CCTTAAAGATGGGCGAGGAAACAGATCCTGTGTGATTTATTTGAAACGGGATTACA 156407  
QY 421 gatttgaatgaagtacccaagttagcattaccattgagaggaacagacagaagaat 480  
|||||  
Db 156408 GATTGTAAATGAAAGTCA-CAAGGTGAGCATTTACCAATGAGAGGAAACAGACGAGAAAT 156466  
QY 481 ctgtatgcttcaacaagacatgcaac 506  
|||||  
Db 156467 CTTGATGGCTTCACAAGACATGCAAC 156492  
RESULT 3  
AL358573/c 267581 bp DNA HTG 15-APR-2001  
LOCUS Homo sapiens chromosome 9 clone RP11-133022, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, 37 unordered pieces.  
ACCESSION AL358573  
VERSION AL358573.17 GI:13660951  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 267581)  
Direct Submission  
Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk

## COMMENT

On Apr 17, 2001 this sequence version replaced g1.13398774.  
----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [humquerry@sanger.ac.uk](mailto:humquerry@sanger.ac.uk)

Project Information

Center project name: BA133022

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 256699 bases at least Q40

Consensus quality: 259744 bases at least Q30

Consensus quality: 261407 bases at least Q20

Insert size: 263981; sum-of-contigs

Insert size: 135491; 19.3% error; agarose-fp

Quality coverage: 5.5ix in Q20 bases; sum-of-contigs quality

coverage: 12.23x in Q20 bases; agarose-fp

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 37 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

* 1 25719 25818: gap of 100 bp
* 25719 25818: contig of 25718 bp in length
* 25819 29567: gap of 100 bp
* 25819 29567: contig of 3749 bp in length
* 29568 29667: gap of 100 bp
* 29568 29667: contig of 10853 bp in length
* 40521 40620: gap of 100 bp
* 40521 40620: contig of 9382 bp in length
* 50003 50102: gap of 100 bp
* 50003 50102: contig of 100 bp
* 50103 54942: gap of 100 bp
* 50103 54942: contig of 4840 bp in length
* 54943 55042: gap of 100 bp
* 54943 55042: contig of 4025 bp in length
* 55068 59167: gap of 100 bp
* 55068 59167: contig of 13160 bp in length
* 59168 72327: gap of 100 bp
* 59168 72327: contig of 13160 bp in length
* 72328 72427: gap of 100 bp
* 72328 72427: contig of 6969 bp in length
* 72428 79397: gap of 100 bp
* 72428 79397: contig of 3242 bp in length
* 79497 82738: gap of 100 bp
* 79497 82738: contig of 2804 bp in length
* 82739 82838: gap of 100 bp
* 82739 82838: contig of 4705 bp in length
* 82839 87643: gap of 100 bp
* 82839 87643: contig of 2204 bp in length
* 87644 89847: gap of 100 bp
* 87644 89847: contig of 2429 bp in length
* 89848 92376: gap of 100 bp
* 89848 92376: contig of 2723 bp in length
* 92377 92476: gap of 100 bp
* 92377 92476: contig of 713 bp in length
* 92477 99800: gap of 100 bp
* 92477 99800: contig of 2218 bp in length
* 99800 102117: gap of 100 bp
* 99800 102117: contig of 2218 bp in length
* 102118 102217: gap of 100 bp
* 102118 102217: contig of 713 bp in length
* 102218 109330: gap of 100 bp
* 102218 109330: contig of 2257 bp in length
* 109331 112187: gap of 100 bp
* 109331 112187: contig of 4120 bp in length
* 112188 112287: gap of 100 bp
* 112188 112287: contig of 2879 bp in length
* 112288 116407: gap of 100 bp
* 112288 116407: contig of 2879 bp in length
* 116408 119386: gap of 100 bp
* 116408 119386: contig of 2879 bp in length
* 119387 119486: gap of 100 bp
* 119387 119486: contig of 2883 bp in length
* 119487 121869: gap of 100 bp
* 119487 121869: contig of 3577 bp in length
* 121870 125446: gap of 100 bp
* 121870 125446: contig of 3344 bp in length
* 125447 128990: gap of 100 bp
* 125447 128990: contig of 4965 bp in length
* 128991 134055: gap of 100 bp
* 128991 134055: contig of 4159 bp in length
* 134056 138314: gap of 100 bp
* 134056 138314: contig of 4159 bp in length

```

```

* 138315 138414: gap of 100 bp
* 138415 140612: contig of 2199 bp in length
* 140613 140712: gap of 100 bp
* 140713 143296: contig of 2584 bp in length
* 143297 143396: gap of 100 bp
* 143397 148113: contig of 4717 bp in length
* 148114 148213: gap of 100 bp
* 148214 151009: contig of 2796 bp in length
* 151010 151109: gap of 100 bp
* 151110 156716: contig of 5507 bp in length
* 156717 158763: gap of 100 bp
* 158764 158863: gap of 100 bp
* 158864 162129: contig of 3266 bp in length
* 162130 162229: gap of 100 bp
* 162230 164443: contig of 2214 bp in length
* 164444 164543: gap of 100 bp
* 164544 169800: contig of 5257 bp in length
* 169801 169900: gap of 100 bp
* 169901 2251545: contig of 81645 bp in length
* 2251546 251645: gap of 100 bp
* 251646 259029: contig of 7384 bp in length
* 259030 259129: gap of 100 bp
* 259130 262871: contig of 3742 bp in length
* 262872 262971: gap of 100 bp
* 262972 265209: contig of 2238 bp in length
* 265210 265309: gap of 100 bp
* 265310 267581: contig of 2272 bp in length.

```

## FEATURES

## source

```

1. .267581
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="9"
  /clone="RP11-133022"
  /clone_1bp="RP11-11.1"
  1. .25718
    /note="assembly_fragment:00645"
    fragment_chain:1
      clone_end:T7
      vector_side:left"
      25819. .29567
        /note="assembly_fragment:01660"
        fragment_chain:1"
      29668. .40520
        /note="assembly_fragment:03080"
        fragment_chain:1"
      40621. .50002
        /note="assembly_fragment:00771"
        fragment_chain:1"
      50103. .54942
        /note="assembly_fragment:02629"
        fragment_chain:1"
      55043. .59067
        /note="assembly_fragment:04565"
        fragment_chain:1"
      59168. .72327
        /note="assembly_fragment:00223"
        fragment_chain:1"
      72428. .79396
        /note="assembly_fragment:03318"
        fragment_chain:1"
      79497. .82738
        /note="assembly_fragment:00224"
        fragment_chain:2"
      82839. .87543
        /note="assembly_fragment:03050"
        fragment_chain:2"
      87644. .89847
        /note="assembly_fragment:01948"
        fragment_chain:3"
      89848. .92376
        /note="assembly_fragment:02191"
        fragment_chain:3"

```

Query Match	97.6%	Score 494	DB 80	Length 267581
Best Local Similarity	99.8%	Pred. No. 1.3e-134		
Matches 505	Conservative	0	Mismatches 0	Indels 1
				Gaps 1
QY 1	caggaagacacaaagaagacacagaggaagtgctcttaataagcactcaattctactca	60		
Db 166149	CAGGAAGACACAAAAGGAAGCACAGAGGTAAGTGCTTTTAAAGCATCTAATTTCTACTCA	166090		
QY 61	gaatttttgatgagccttaagttccctactcgtttctatacctctactcaactcgtcc	120		
Db 166089	GAAATTTTGTATGGCCCTTAATCTCTACTACGTTTCTATCTCTTCCATCTACTCTCCTC	166030		
QY 121	cgggaatccactacacgaattttctattttcttgccctgattgtttctgacgagcccaattgga	180		
Db 166029	CCGGAATCCACTACCGAATTTCTATTTCTTCTCTCTCTGTAATGCTATCGGCTCACTTGGA	165970		
QY 181	tttaacctcacagagctcgtgaatttctcaaccggagctcaacctccgctccctccatattg	240		
Db 165969	TTTATCCACAGAGAGTGCGATTTTCTTATCCGGGGGTACCTCGTCCCTCATTTTGTG	165910		
QY 241	ctccacttcaacagatccctcgtggagaaatgcgcggcgccacatttggatccatgatgagc	300		
Db 165909	CTCCACTTTTCACAATCCCTCTGGAGAAATGCCCGGCCGCATCTTGGGTCATCGATGAGC	165850		
QY 301	ctggcccccgtgcctcgttcgccgctgttggagggaagaacattgaaataatgatggtgt	360		
Db 165849	CTGCCCCCTGTGCTCGTGTCGCCCTGTGTGAGGGAAGGACATTTGAAAATGATGATGATGT	165790		
QY 361	ccttaaaagatcgtggcaggaacacagatccctgttctgataattatttgaacgaggaattaca	420		
Db 165789	CCTTAAAGGATGGGCAGGAAAACAAGATCTGTGTGGATATTATTATTGAACGGATTTACA	165730		
QY 421	gatttgaatggaagtccaccaagtggagcattaccaatgagaggaacacagacgagaaat	480		

QY	481	ctgatggtcttacaagaacatgcac 506	
Db	165670	CTTGATGGCTTCACAAAGCATGCAAC	165645
RESULT	4		
LOCUS	AF103908	5435 bp DNA	PRI 14-AUG-2000
DEFINITION	Homo sapiens non-coding RNA	DD3 gene, exons 2, 3, and 4.	
ACCESSION	AF103908		
VERSION	AF103908.1	GI:6165974	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.		
AUTHORS	1 (bases 1 to 5435) Bussemakers, M.J., van Bokhoven, A., Verhaegh, G.W., Smit, F.P., Karihaas, H.F., Schalken, J.A., Debruyne, F.M., Ru, N. and Isaacs, W.B. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer		
TITLE	Cancer Res. 59 (23), 5975-5979 (1999)		
JOURNAL	20072260		
MEDLINE	10606244		
PUBMED	2 (bases 1 to 5435)		
REFERENCE	Bussemakers, M.J.G., Van Bokhoven, A., Verhaegh, G.W., Smit, F.P., Karihaas, H.F.M., Schalken, J.A., Debruyne, F.M.J., Ru, N. and Isaacs, W.B.		
AUTHORS	Direct Submission Submitted (28-OCT-1998) Urology Research Laboratory, University Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands		
TITLE	Location/Qualifiers		
JOURNAL	1. 5435		
FEATURES	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="9" /map="9q21-q22" 1. .78 /rpt_family="Alu" /rpt_type="dispersed 533. .697 /note="alternative exon present in 5% of cDNA clones" /number=2 1035. .1294 /rpt_family="Alu" /rpt_type="dispersed join(1571. .1753,1981. .5435) /note="transcript II" /product="non-coding RNA DD3" join(1571. .1753,1981. .2517) /note="transcript I" /product="non-coding RNA DD3" join(1571. .1753,1981. .3579) /note="transcript (major) II" 5268. .5423 /note="LINE" /rpt_family="L1" /rpt_type="dispersed		

polyA_site	5435	/note="transcript III"	
BASE COUNT	1546 a	1188 c	1036 g 1654 t
ORIGIN			11 others

Query Match	95.30;	Score 482;	DB 86;	Length 5435;
Best Local Similarity	99.60;	Pred. No. 2.6e-131;		
Matches 504;	Conservative 0;	Mismatches 0;	Indels 2;	Gaps 2

QY	1	caagaaagcaaaagaagaaagacagaggaagaaagcttaagagcttaaaagcaactaaattcacca	60
Db	1728	CAGAAACCAAAAAGGAGACAGACAGAGCTAAGTGCCTTTATAAAGCACTCAATTTCTACTCA	1787
QY	61	gaaaatttcgagcgcttlaagcttcctctaaactcgcgtttctatcctctcaactcagtgctc	120
Db	1788	GAATTTTATATGGCCCTTAAGTTCCTCTACCTGTTTCTATCCCTTACTCACTGTCCTC	1847
QY	121	ccggaatccacacacagatatttcattcttcctgcctcgtatctctgaacccgagccacttga	180
Db	1848	CCGGAATCCACTACCGATTTTCTATTCTTTGCTCTGCTATATGTCATAGCAGCAGCTTGA	1907
QY	181	tttaatccacagagagcttgatatttcaaccgagccacactccgtccctccatatttgc	240
Db	1908	TTTTAT-CTCAGGAGTCTGGATTTTCTACCCGAGGCTACCTCCGTCCTCCATATTGTGC	1966
QY	241	ctcgaactttacagatattccctcggaggaagaatgcccgcgcgcacattcggatcatcgaatgc	300
Db	1967	CTCCACTTTTACAAATATCCCTGGGAGAAATGCCCGCGCATCTTGGGTCATCGATTGAGC	2028
QY	301	ctcgccctcgagccctggctcccgctctgaggaagaagacattagaataatgaattgattgtc	360
Db	2027	CTCGCCCTTGSCCTGGGCCGCTTGGAGGAGGAAGCATTAGAAAATGAAATTGATGCTTT	2086
QY	361	ccctaaagaagtgaggccggaataaagatcctctgttggataattatttgaacgggaattca	420
Db	2087	CCTTAAAGATGGGCGAGGAATACAGATCCGTGTGGATATTATTATTTGAACGGATTGCA	2146
QY	421	gatttgaagaagagtcacacaaatgagcatctcaaatgagaggaagaacagacagagaatat	480
Db	2147	GATTGGAATATGAAGTCA-CAAACTGAGCATTTACATATGAGAGGAATACAGACGAATAAT	2205
QY	481	cttgatgagcttccacaaagacatgtaac	506
Db	2206	CTTGATGAGCTTCCAAAGACATGTGAAC	2231

RESULT	5
LOCUS	AF103907
DEFINITION	AF103907 3923 bp mRNA PRI 14-AUG-2000
ACCESSION	AF103907
VERSION	AF103907.1
KEYWORDS	GI:6165973
SOURCE	.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE	1 (bases 1 to 3923)
JOURNAL	Bussenmakers, M.J.T., Van Bokhoven, A., Verhaegh, G.W., Smits, F.P., Kerkhofs, H.F., Schalken, J.A., Debruyne, F.M., Ru, N. and Isaacs, W.B.
MEDLINE	DD3: a new prostate-specific gene, highly overexpressed in prostate cancer
PUBMED	Cancer Res. 59 (23), 5975-5979 (1999)
REFERENCE	20072260
AUTHORS	10606244
	2 (bases 1 to 3923)
	Bussenmakers, M.J.T.G., Van Bokhoven, A., Verhaegh, G.W., Smits, F.P., Kerkhofs, H.F.M., Schalken, J.A., Debruyne, F.M.T., Ru, N. and Isaacs, W.B.
TITLE	Direct Submission
JOURNAL	Submitted (28-OCT-1998) Urology Research Laboratory, University Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands

FEATURES	Location/Qualifiers
source	1. .3923

repeat\_region

```

polyA_site      /rpt_type=dispersed
BASE COUNT      1157 a      840 c      740 g      1186 t
ORIGIN

```

Query Match	48.1%;	Score 243.4;	DB 88;	Length 3923;
Best Local Similarity	99.2%;	Pred. No. 6.6e-61;		
Matches 255;	Conservative 0;	Mismatches 1;	Indels 1;	Gaps 1;

QY	250	caagatccctcctggagaaagaaibcccgccgcgcacatcttggatcatalcatgaatgaacccgcgcctg	309
Db	464	CAGAGATCCCTGGAGAAATCCCGCGCCGCATCTTGGATCATCGATGAGACCTTCGCCCTG	5233
QY	310	tgacctgtgctccgcctctgtgaagaaagacatbagaanaatgaatgatgtgttccttcaag	369
Db	524	TGCTGTGGTCCGCTTGTGAGGGAAGACATAGAAATGATGATGTGTCTCTTAAAGG	5833
QY	370	atggggcaggaaaacagatccctggtgtgatattatttgaagggaattaaagttgaaa	429
Db	584	ATGGGCAAGAAAACAGATCCGTGTGTGATATTATTATTTGAACGGGATTAACGATTTGAAA	643
QY	430	tgaagtcacccaagtgtgacatctaaccaatgaagagaaaacagacagagaaaaatcttgtatggc	489
Db	644	TGAAGTCA - CAAAGTGAAGCATATTACATTAAGAGAGAAAACAGAGAAAATCTTGATGGC	702
QY	490	ttcccaagacatgtcaac	506
Db	703	TTTCAAGACATGCAAC	719

FEATURES	source
REFERENCE	1 (bases 1 to 7218)
AUTHORS	Donner, F., Schefflilinger, F. and Palkner, F. Gunter.
TITLE	Recombinant fowlpox virus
JOURNAL	Patent: US 5670367-A 14 23-SEP-1997;
LOCUS	166494
DEFINITION	Sequence 14 from patent US 5670367.
ACCESSION	166494
VERSION	166494.1 GI:2724471
KEYWORDS	Unknown.
SOURCE	Unknown.
ORGANISM	Unclassified.
PART	7218 bp DNA

	Query Match	11.5%	Score 58.2	DB 10;	Length 7218;
	Best Local Similarity	2.1%;	Pred. No. 3.3e-06;		
	Matches	6;	Conservative 182;	Mismatches 95;	Indels 0;
	Gaps	0;			
Oy	42	agaactcaattctcactagaaattttgtaggcgttaagtccctactcgtttctatc	101		
		:	:	:	:
Dd	1057	AGCTTGCATT	1116		
OY	102	ccttcctactacgylcccgcgaatccactacgaaatttcatlcttgctcgatlty	161		
	:::::	: : : : : :	:	:	:

Db	1117	XXX	XXX	XXX	XXX
Oy	162	tctgactgcgtcacttgattatccacaggaabcltggattttctaccggcgctacct	221		
Db	1177	XXX	XXX	XXX	XXX
Oy	222	cgcgcccccacatttgcctcccaacttcacaagatccctggaggagaatgccgcgcaca	281		
Db	1237	XXX	XXX	XXX	XXX
Oy	282	tcttgagtcacgatagagccgcgcctgctgctgctcgcgcctt	324		
Db	1297	XXX	XXX	XXX	XXX
RESULT	7				
LOCUS	AX018075	359 bp	DNA	PAT	07-SEP-2000
DEFINITION	Sequence 23 from Patent WO946374.				
ACCESSION	AX018075				
VERSION	AX018075.1	GI:10042526			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
TITLE	1 (bases 1 to 359)				
JOURNAL	Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and Plarsky,C.				
FEATURES	Human nucleic acid sequences from prostate tumour tissue				
SOURCE	Patent: WO 946374-A 23-16-SEP-1999; SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN BERND (DE); ROSENTHAL ANDRE (DE); METAGSKY GES FUER GENOMFORSCHUN (DE); PILARSKY CHRISTIAN (DE) Location/Qualifiers 1..359				
BASE COUNT	121 a	75 c	94 g	69 t	
ORIGIN	/organism="Homo sapiens" /db_xref="taxon:9606"				
Query Match	10.4%; Score 52.4; DB 9; Length 359;				
Best Local Similarity	98.1%; Pred. No. 0.00011;				
Matches	53; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
Oy	250	cacgagatccccggaggaaatcccgcgccgcatcttggtcatogatagctc	303		
Db	306	CAGAGATCCCTGGGAGAANAATCCCGGCCGCATCTTGSGGTATCATGATGACTC	359		
RESULT	8				
LOCUS	HS30M3	102200 bp	DNA	PRT	23-NOV-1999
DEFINITION	Human DNA sequence from clone 30M3 on chromosome 6p22.1-22.3. Contains three novel genes, one similar to C. elegans Y63D3A.4 and one similar to (predicted) plant, worm, yeast and archaea bacterial genes, and the first exon of the KIAA0319 gene. Contains ESTs, GSSS and putative CpG islands, complete sequence.				
ACCESSION	AL031775				
VERSION	AL031775.1	GI:4071041			
KEYWORDS	HTG; CpG island; KIAA0319; Y63D3A.4.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
TITLE	1 (bases 1 to 102200)				
JOURNAL	Phillips,S. Direct Submission Submitted (05-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerquest@sanger.ac.uk On Dec 29, 1998 this sequence version replaced gi:4007165.				
COMMENT					

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence is the entire insert of clone 30M3. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>

30M3 is from the library RPCI1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcrPAC2>.

Location/Qualifiers  
1..102200

```

HMS
source
Location/Qualifiers
1..102200
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="p22.1-22.3"
/clone="RP1-30M3"
/clone_11b="RP1-1"
757..825
/note="MER47 repeat: matches 9..76 of consensus"
829..932
/note="MER47 repeat: matches 2226..2319 of consensus"
905..976
/note="MER47 repeat: matches 2249..2323 of consensus"
1320..1641
/note="AluSc repeat: matches 1..299 of consensus"
1642..1927
/note="AluYb repeat: matches 1..289 of consensus"
2392..2517
/note="FLAM_C repeat: matches 1..126 of consensus"
2760..2789
/note="15 copies 2 mer tg 97% conserved"
3013..3313
/note="AluJo repeat: matches 1..301 of consensus"
3467..3763
/note="AluSp repeat: matches 1..229 of consensus"
complement(3650..4203)
/note="GSS B63434"
4458..4591
/note="MIR repeat: matches 53..208 of consensus"
4799..4952
/note="MER3 repeat: matches 3..154 of consensus"
join(<5119..5253,7768..9169)
/gene="dJ30M3.2"
AA259531 AA289775 N50863 AA806198
AA931087 AA830274 AA199746 AA912273 AT081164 AT88433
AA604069 AA772465 AA876636 AA525057 AA147796 AA639167
AA656292 AA289775 AA259531 AA731055 AA290469 D20450
AA517053 AA669103 AA974142 AA903560 AT083595 AT287615
AT041875 AT143439 N45220 R23494 AA798467 AA798123 AA51228
AA473082 AT120905 AA663774 AA840219 AA041897 D77719
AT019712 AT006885 AT194277 D56188 AA926179 AA851214 T51937
/evidence-not_experimental
/product="dJ30M3.2 (novel protein)"
5119..9169
/gene="dJ30M3.2"
join(<5119..5253,7768..7893)
/gene="dJ30M3.2"
/note="supported by GENSCAN"
/codon_start=1
/evidence-not_experimental
/product="dJ30M3.2 (novel protein)"

```

```

/protein_id="CAA21140.1"
/db_xref="GI:4210352"
/translation="AKFEFHGDYKOPHLVSRDKTGIVVNNQSVFLEIDRQHL
QTPKMKATIFKCSICLTLPQQLTHMAGTIEDLRPEMP"
repeat_region
  /note="AluSg repeat: matches 1. .291 of consensus"
  5393. .5674
repeat_region
  /note="AluSg repeat: matches 1. .49 of consensus"
  5715. .5763
repeat_region
  /note="MER5a repeat: matches 1. .189 of consensus"
  5883. 6000
repeat_region
  /note="AluY repeat: matches 1. .295 of consensus"
  6010. 6308
repeat_region
  /note="AluY repeat: matches 1. .296 of consensus"
  6384. .6667
repeat_region
  /note="AluJo repeat: matches 5. .207 of consensus"
  6740. .6878
repeat_region
  /note="MIR repeat: matches 69. .207 of consensus"
  7106. .7400
repeat_region
  /note="AluJb repeat: matches 1. .297 of consensus"
  8676. .8721
repeat_region
  /note="23 copies 2 mer tt 78% conserved"
  9092. .9129
repeat_region
  /note="19 copies 2 mer aa 82% conserved"
  9124. .9129
polyA_signal
  /gene="dJ30M3.2"
  9169
polyA_site
  /gene="dJ30M3.2"
  9439. .9621
repeat_region
  /note="MIR repeat: matches 8. .256 of consensus"
  9691. .9798
repeat_region
  /note="L1MC4 repeat: matches 7851. .7977 of consensus"
  10020. .10149
repeat_region
  /note="L1MD3 repeat: matches 7361. .7499 of consensus"
  10507. .10716
repeat_region
  /note="L1MC4 repeat: matches 6633. .6858 of consensus"
  10855. .11115
repeat_region
  /note="AluJo repeat: matches 1. .283 of consensus"
  11469. .11619
repeat_region
  /note="MIR1A1 repeat: matches 1. .146 of consensus"
  11620. .11806
repeat_region
  /note="AluSx repeat: matches 1. .287 of consensus"
  11907. .12055
repeat_region
  /note="MIR1A1 repeat: matches 146. .310 of consensus"
  complement(join(12316. .12571,15963. .>16147))
  /gene="dJ30M3.1"
  /note="mach: ESTs AA5222110 AA207689 AA819653 AA928745
  AI017897 AA993303 AI216348 AA832106 AI213362 AA552456
  AA628986 AA614605 T55234 AA121362 F17737 AA445103 AA073807
  AA121372 AI052396 AA939005 AI267851 T53974; supported by
  GENSCAN and FGENES"
  /evidence="not_experimental"
  /product="dJ30M3.1 (novel protein similar to (predicted)
  plant, worm, yeast and archaea bacterial proteins)"
  complement(join(12316)
  /gene="dJ30M3.1"
  complement(12316. .16147)
  /gene="dJ30M3.1"
  complement(12331. .12336)
  /gene="dJ30M3.1"
  complement(join(12415. .12571,15963. .>16147))
  /gene="dJ30M3.1"
  /note="match: proteins P87304 P93828 CE02501 CE00158
  P95914; supported by GENSCAN and FGENES"
  /codon_start=1
  /evidence="not_experimental"
  /product="dJ30M3.1 (novel protein similar to (predicted)
  plant, worm, yeast and archaea bacterial proteins)"
  /protein_id="CAA21139.1"
  /db_xref="GI:4210351"
  /db_xref="GI:4210351"
  /db_xref="SPTREMBL:095549"
  /translation="ITLVSAAPGVICEMKVEEHTNAGTGLGATATLVDNISTMA
  LTRGAPGVSYDMNITMSPAKLGEIDIVITAHVLAOGKTLAFTSDVLTAKAKSKLI
  AOGHRTKHLGN"

```

```

repeat_region
  13295. .13604
  /note="AluSx repeat: matches 1. .310 of consensus"
  14521. .14821
  /note="AluYb repeat: matches 1. .299 of consensus"
  15123. .15387
  /note="AluSx repeat: matches 30. .298 of consensus"
  15514. .15624
  /note="AluJo repeat: matches 2. .136 of consensus"
  16957. .17250
  /note="AluSg repeat: matches 1. .294 of consensus"
  17526. .17589
  /note="MIR repeat: matches 194. .258 of consensus"
  17534. .17637
  /note="L2 repeat: matches 2588. .2692 of consensus"
  17943. .18245
  /note="AluYas repeat: matches 1. .311 of consensus"
  18551. .18684
  /note="L2 repeat: matches 2357. .2495 of consensus"
  18704. .18998
  /note="AluSx repeat: matches 6. .300 of consensus"
  19037. .19404
  /note="match: GSS B40022"
  19504. .19639
  /note="MIR repeat: matches 88. .238 of consensus"
  20030. .20303
  /note="AluSx repeat: matches 24. .297 of consensus"
  complement(20331. .20463)
  /note="match: GSS B74972"
  21207. .21337
  /note="MIR repeat: matches 90. .244 of consensus"
  21347. .21600
  /note="AluJo repeat: matches 36. .287 of consensus"
  21692. .21891
  /note="L2 repeat: matches 1320. .1536 of consensus"

Query Match
Best Local Similarity 8.5%; Score 43; DB 92; Length 102200;
Matches 94; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Qy 66 ttltgatgccttaagttcctactcgtctatcctcctcactcactgctccgga 125
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71711 TTAAATGTCCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCC 71770

Qy 126 atccactacgagttcttctattctctgctcgtatgctgactgctcactgattat 185
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71771 CCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTC 71830

Qy 186 cctcagcagagctgatttctcaccgagctcactcgtcctccatattgctctcc 244
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71831 TTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTCC 71889

RESULT 9
AC058821/c DNA HTG 24-FEB-2001
LOCUS AC058821/c Homo sapiens chromosome 4 clone RP11-778B12, WORKING DRAFT
DEFINITION SEQUENCE, 46 unordered pieces.
ACCESSION AC058821
VERSION AC058821.2 GI:7767912
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 215103)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 215103)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

```



## COMMENT

MO 63108, USA  
On May 10, 2000 this sequence version replaced gi:7596990.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H\_NH0778812  
----- Summary Statistics -----  
Sequencing vector: M13, 1008  
Sequencing vector: plasmid: 08  
Chemistry: Dye-terminator Big Dye, 08 of reads  
Chemistry: Dye-terminator Big Dye, 08 of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 189194 bases at least Q40  
Consensus quality: 197110 bases at least Q30  
Consensus quality: 201125 bases at least Q20  
Insert size: 207000; agarose-fp  
Insert size: 210603; sum-of-ctrls  
Quality coverage: 3.08 in Q20 bases; agarose-fp  
Quality coverage: 3.01 in Q20 bases; sum-of-ctrls  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 46 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of 'N', but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1190: contig of 1190 bp in length  
\* 1191 1280: gap of unknown length  
\* 1291 2491: contig of 1201 bp in length  
\* 2492 2591: gap of unknown length  
\* 2592 3970: contig of 1379 bp in length  
\* 3971 4070: gap of unknown length  
\* 4071 6259: contig of 2189 bp in length  
\* 6260 6359: gap of unknown length  
\* 6360 8648: contig of 2289 bp in length  
\* 8649 8749 10583: contig of 1835 bp in length  
\* 8749 10583: contig of 1835 bp in length  
\* 10584 10683: gap of unknown length  
\* 10684 13046: contig of 2363 bp in length  
\* 13047 13146: gap of unknown length  
\* 13147 15724: contig of 2578 bp in length  
\* 15725 15824: gap of unknown length  
\* 15825 18468: contig of 2644 bp in length  
\* 18469 18568: gap of unknown length  
\* 18569 21077: contig of 2509 bp in length  
\* 21078 21177: gap of unknown length  
\* 21178 24304: contig of 3127 bp in length  
\* 24305 24404: gap of unknown length  
\* 24405 27866: contig of 3462 bp in length  
\* 27867 27966: gap of unknown length  
\* 27967 31186: contig of 3220 bp in length  
\* 31187 31286: gap of unknown length  
\* 31287 34406: contig of 3120 bp in length  
\* 34407 34506: gap of unknown length  
\* 34507 36959: contig of 2453 bp in length  
\* 36960 37059: gap of unknown length  
\* 37060 40366: contig of 3207 bp in length  
\* 40367 40366: gap of unknown length  
\* 40367 43164: contig of 2798 bp in length  
\* 43165 43264: gap of unknown length  
\* 43265 46349: contig of 3085 bp in length  
\* 46350 46449: gap of unknown length  
\* 46450 49752: contig of 3303 bp in length  
\* 49753 49852: gap of unknown length  
\* 49853 52949: contig of 3097 bp in length  
\* 52950 55737: contig of 2688 bp in length  
\* 55738 55837: gap of unknown length

## FEATURES

## source

55838 58954: contig of 3117 bp in length  
\* 58955 59054: gap of unknown length  
\* 59055 62369: contig of 3315 bp in length  
\* 62370 62469: gap of unknown length  
\* 62470 65962: contig of 3493 bp in length  
\* 65963 66062: gap of unknown length  
\* 66063 70029: contig of 3967 bp in length  
\* 70030 70129: gap of unknown length  
\* 70130 73918: contig of 3789 bp in length  
\* 73919 74018: gap of unknown length  
\* 74019 78049: contig of 4031 bp in length  
\* 78050 78149: gap of unknown length  
\* 78150 84568: contig of 6419 bp in length  
\* 84569 84668: gap of unknown length  
\* 84669 91351: contig of 6683 bp in length  
\* 91352 91451: gap of unknown length  
\* 91452 96851: contig of 5400 bp in length  
\* 96852 96951: gap of unknown length  
\* 96952 102096: contig of 5145 bp in length  
\* 102097 102196: gap of unknown length  
\* 102197 107711: contig of 5515 bp in length  
\* 107712 107811: gap of unknown length  
\* 107812 111824: contig of 4013 bp in length  
\* 111825 111924: gap of unknown length  
\* 111925 116715: contig of 4791 bp in length  
\* 116716 116815: gap of unknown length  
\* 116816 121462: contig of 4647 bp in length  
\* 121463 121562: gap of unknown length  
\* 121563 127640: contig of 6078 bp in length  
\* 127641 127740: gap of unknown length  
\* 127741 133093: contig of 5353 bp in length  
\* 133094 133193: gap of unknown length  
\* 133194 139965: contig of 6772 bp in length  
\* 139966 140065: gap of unknown length  
\* 140066 146801: contig of 6736 bp in length  
\* 146802 146901: gap of unknown length  
\* 146902 154099: contig of 7198 bp in length  
\* 154100 154199: gap of unknown length  
\* 154200 163123: contig of 8924 bp in length  
\* 163124 163223: gap of unknown length  
\* 163224 171228: contig of 8005 bp in length  
\* 171229 171328: gap of unknown length  
\* 171329 179017: contig of 7689 bp in length  
\* 179018 179117: gap of unknown length  
\* 179118 188418: contig of 9301 bp in length  
\* 188419 188518: gap of unknown length  
\* 188519 199297: contig of 10779 bp in length  
\* 199298 199397: gap of unknown length  
\* 199398 215103: contig of 15706 bp in length.  
Location/Qualifiers  
1. 215103  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/clone="RP11-778B12"  
1. 1190  
/note="assembly\_name:Contig13"  
1291. 2491  
/note="assembly\_name:Contig18"  
2592. 3970  
/note="assembly\_name:Contig19"  
4071. 6259  
/note="assembly\_name:Contig20"  
6360. 8648  
/note="assembly\_name:Contig21"  
8749. 10583  
/note="assembly\_name:Contig22"  
10684. 13046  
/note="assembly\_name:Contig23"  
13147. 15724  
/note="assembly\_name:Contig25"  
15825. 18468  
/note="assembly\_name:Contig26"







chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Ch1>  
RP4-781K5 is from the library RPI-4 constructed by the group of  
Pleier de Jong. For further details see  
<http://www.chori.org/Dacpac/home.htm>

## VECTOR: pCYPAC2

This sequence is the entire insert of clone RP4-781K5 The true left  
end of clone RPI1-400F1 is at 75536 in this sequence. The true left  
end of clone RPI1-380I8 is at 104240 in this sequence. The true  
right end of clone RP5-855F14 is at 15526 in this sequence.

## FEATURES

```
source
1..132887
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q42.1-43"
/clone="RP4-781K5"
/clone.lib="RPI-4"
3739..6486
/note="Cpg island"
/evidence="not_experimental"
repeat_region
4673..4765
/note="31 copies 3 mer ggc 65% conserved"
4752..4817
/note="33 copies 2 mer cc 68% conserved"
4816..5031
/note="72 copies 3 mer ggc 59% conserved"
4951..5248
/note="149 copies 2 mer gg 58% conserved"
4961..5235
/note="5 copies 55 mer 65% conserved"
5125..5424
/note="Single clone region. Sequence from reads from a  
short insert library derived from a single m13 clone.  
Restriction digest data confirm the assembly."
5333..5374
/note="14 copies 3 mer ccg 81% conserved"
10072..10367
/note="Alusg repeat: matches 1..296 of consensus"
11031..11339
/note="Alusx repeat: matches 1..308 of consensus"
12204..12235
/note="8 copies 4 mer tacc 90% conserved"
12993..13160
/note="84 copies 2 mer at 75% conserved"
13947..13982
/note="18 copies 2 mer ac 83% conserved"
15995..16181
/note="Aluud repeat: matches 1..167 of consensus"
16181..16389
/note="Aluub repeat: matches 112..310 of consensus"
17813..18003
/note="Tiggr2a repeat: matches 1..191 of consensus"
18004..18297
/note="Alusg repeat: matches 1..306 of consensus"
18298..18539
/note="Tiggr2a repeat: matches 190..434 of consensus"
19058..19133
/note="19 copies 4 mer acac 72% conserved"
19686..19940
/note="MIR repeat: matches 2..261 of consensus"
20066..20186
/note="L2 repeat: matches 1304..1432 of consensus"
20319..20364
/note="L2 repeat: matches 2699..2744 of consensus"
20532..20840
/note="Aluub repeat: matches 1..307 of consensus"
21075..21236
/note="FRAM repeat: matches 4..165 of consensus"
21242..21341
/note="Alusg repeat: matches 5..304 of consensus"
21820..21882
/note="ORSL repeat: matches 369..427 of consensus"
21883..22200
/note="Alub repeat: matches 1..312 of consensus"
22416..22682
/note="L2 repeat: matches 1572..1842 of consensus"
23587..23754
/note="MIR repeat: matches 3..172 of consensus"
24665..25041
/note="MIR1 repeat: matches 109..513 of consensus"
25058..25435
/note="MIR1 repeat: matches 101..471 of consensus"
25452..25623
/note="L2 repeat: matches 2113..2285 of consensus"
25658..25932
/note="L2 repeat: matches 2424..2731 of consensus"
26011..26079
/note="TIGGR2 repeat: matches 2633..2707 of consensus"
26377..26678
/note="AluY repeat: matches 1..302 of consensus"
27536..27844
/note="Aluud repeat: matches 1..302 of consensus"
29258..29705
/note="MIR1C repeat: matches 1..466 of consensus"
30014..30311
/note="Alusg repeat: matches 1..298 of consensus"
32010..32165
/note="LIPB3 repeat: matches 5901..6024 of consensus"
32166..32468
/note="Alusx repeat: matches 1..303 of consensus"
32469..32591
/note="LIPB3 repeat: matches 6024..6150 of consensus"
33112..33418
/note="AlusC repeat: matches 1..306 of consensus"
34819..34910
/note="MIR repeat: matches 41..131 of consensus"
35691..36075
/note="7 copies 55 mer 64% conserved"
35692..36075
/note="192 copies 2 mer aa 60% conserved"
35914..36033
/note="30 copies 4 mer aaag 75% conserved"
36279..36342
/note="16 copies 4 mer agag 76% conserved"
37040..37216
/note="LIPB6 repeat: matches 5308..5482 of consensus"
37217..37510
/note="Alusg repeat: matches 1..294 of consensus"
37511..38166
/note="LIPB6 repeat: matches 5482..6146 of consensus"
38996..39207
/note="MIR repeat: matches 46..251 of consensus"
39330..39714
/note="MIR1B repeat: matches 1..390 of consensus"
40020..40075
/note="14 copies 4 mer tcta 71% conserved"
41109..41437
/note="MIR1 repeat: matches 66..547 of consensus"
42960..43107
/note="L2 repeat: matches 2605..2748 of consensus"
43297..43583
/note="Alusg repeat: matches 11..297 of consensus"
43628..43713
/note="MIR repeat: matches 158..262 of consensus"
43714..44269
/note="MIR1F repeat: matches 1..541 of consensus"
44270..44325
/note="MIR repeat: matches 108..158 of consensus"
44618..44680
/note="MIR2D repeat: matches 489..552 of consensus"
44719..45218
/note="MIR2D repeat: matches 1..533 of consensus"
45344..45371
/note="14 copies 2 mer tg 96% conserved"
```



*	1050.0	113299:	contig of 7866 bp in length
*	113396	113399:	gap of unknown length
*	113396	123411:	contig of 10016 bp in length
*	123412	123511:	gap of unknown length
*	123512	124559:	contig of 11048 bp in length
*	134560	134659:	gap of unknown length
*	134660	145428:	contig of 10766 bp in length
*	145429	145528:	gap of unknown length
*	145529	157178:	contig of 11650 bp in length
*	157179	157278:	gap of unknown length
*	157279	168351:	contig of 11073 bp in length
*	168352	168451:	gap of unknown length
*	168452	183191:	contig of 14740 bp in length
*	183192	183291:	gap of unknown length
*	183292	199103:	contig of 15812 bp in length

URES	Location/Qualifiers
source	1. 199103
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="OL"
	/clone="RP11-424d2"
misc_feature	1. 1409
	/note="assembly_name:Contig20"
misc_feature	1510. 3167
	/note="assembly_name:Contig21"
misc_feature	3268. 4730
	/note="assembly_name:Contig22"
misc_feature	4831. 6594
	/note="assembly_name:Contig23"
misc_feature	6695. 8447
	/note="assembly_name:Contig24"
misc_feature	8548. 11042
	/note="assembly_name:Contig25"
misc_feature	11143. 13409
	/note="assembly_name:Contig26"
misc_feature	13510. 15867
	/note="assembly_name:Contig27"
misc_feature	15888. 17692
	/note="assembly_name:Contig28"
misc_feature	17793. 19183
	/note="assembly_name:Contig29"
misc_feature	19284. 21394
	/note="assembly_name:Contig30"
misc_feature	21495. 23887
	/note="assembly_name:Contig31"
misc_feature	23988. 25661
	/note="assembly_name:Contig32"
misc_feature	25762. 28744
	/note="assembly_name:Contig33"
misc_feature	28845. 32323
	/note="assembly_name:Contig34"
misc_feature	32424. 35441
	/note="assembly_name:Contig35"
misc_feature	35942. 38154
	/note="assembly_name:Contig36
	clone.end:SP6
misc_feature	vector.side:left"
	38255. 38796
misc_feature	/note="assembly_name:Contig7"
	38887. 42479
misc_feature	/note="assembly_name:Contig37"
	42580. 45860
misc_feature	/note="assembly_name:Contig38"
	45961. 49485
misc_feature	/note="assembly_name:Contig39"
	49566. 54307
misc_feature	/note="assembly_name:Contig40"
	55008. 60322
misc_feature	/note="assembly_name:Contig41"
	60423. 65547
misc_feature	/note="assembly_name:Contig42"
	65348. 71207
misc_feature	/note="assembly_name:Contig43"

misc_feature	71308. .78131
/note="assembly_name:Contig44"	
misc_feature	78232. .87106
/note="assembly_name:Contig45"	
misc_feature	87207. .96785
/note="assembly_name:Contig46"	
misc_feature	96886. .105329
/note="assembly_name:Contig47"	
misc_feature	105430. .113295
/note="assembly_name:Contig48"	
misc_feature	111396. .123411
/note="assembly_name:Contig49"	
misc_feature	123512. .134559
/note="assembly_name:Contig50"	
misc_feature	134660. .145428
/note="assembly_name:Contig51"	
misc_feature	145529. .157178
/note="assembly_name:Contig52"	
misc_feature	157279. .168351

	Query Match	7.9%	Score 40;	DB 76;	Length 199103;	
	Best Local Similarity	51.1%;	Pred. No. 1.2;			
	Matches 94;	Conservative	0;	Mismatches	90;	Indels 0; Gaps 0;
QY	258	cctggagaagaattcccgagcgcgaccatcttggtgcacgatgaagccctgtgcctg	317			
Db	189808	CCTGGGACCANTGTGGGGCTAAATGGTGAGTCACAGAAGCGTCCAGCTGGGCCAGGG	189867			
OY	318	cccgttctggagggaagacattagaanaatgatgtgttccttaagatgtgcag	377			
Db	189868	CTATCAGAGAGCTGCATTAAGAACCTTTACTCCTGATGAGTAGTGTGGACAGCGCAATCAG	189927			
OY	378	gaacaacagtctgtgtgtgtgataattttttaagsggatcaagtttgaatgaatca	437			
Db	189928	GGAATCAAAGGCGTGAAGGTGTATTATCATACAGCCAGCGCTCTCAATCAAAAAATAATAAA	189987			
OY	438	ccaa 441				
Db	189988	TAAA 189991				
RESULT	14					
LOCUS	AC005225	166774 bp	DNA			
DEFINITION	Homo sapiens PAC clone RP3-414A15 from 14q24.3, complete sequence.		PRI			30-SEP-2000
ACCESSION	AC005225					
VERSION	AC005225.2	GI:6094661				
KEYWORDS	HTG.					
SOURCE	human.					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
REFERENCE						
AUTHORS	1 (bases 1 to 166774)					
TITLE	Sulston,J.E. and Waterston,R.					
JOURNAL	Toward a complete human genome sequence					
REFERENCE	Genome Res. 8 (11), 1097-1108 (1998)					
MEDLINE	99063792					
AUTHORS	2 (bases 1 to 166774)					
TITLE	All,J., Wohlman,P. and Elliott,G.					
JOURNAL	The sequence of Homo sapiens PAC clone RP3-414A15					
REFERENCE	Unpublished					
AUTHORS	3 (bases 1 to 166774)					
TITLE	Waterston,R.H.					
JOURNAL	Direct Submission					
	Submitted (01-JUL-1998) Genome Sequencing Center, Washington					
	University School of Medicine, 4444 Forest Park Parkway, St. Louis,					
	MO 63108, USA					
REFERENCE	4 (bases 1 to 166774)					
AUTHORS	Waterston,R.H.					
TITLE	Direct Submission					
JOURNAL	Submitted (22-OCT-1999) Genome Sequencing Center, Washington					
	University School of Medicine, 4444 Forest Park Parkway, St. Louis,					
	MO 63108, USA					

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

5 (bases 1 to 166774)  
Waterston, R.  
Direct Submission  
Submitted (30-SEP-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Oct 22, 1999 this sequence version replaced gi:3450912.

## COMMENT

-----  
Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: MUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: saplensew@wustl.edu  
-----  
Summary Statistics  
Center project name: H\_DJ0414A15  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:  
This clone from chromosome 14 was provided by Dr. Pieter de Jong,  
Roswell Park Cancer Institute, Human Genetics Department, Elm and  
Carlton Streets, Buffalo NY 14263-0001 USA.

SOURCE INFORMATION:  
This clone was derived from human PAC library RPT-3, prepared by  
Pieter de Jong and coworkers at the Roswell Park Cancer Institute  
(<http://hscpac.med.buffalo.edu>) using the method described by  
Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from  
one male donor.  
The clone may be obtained either from Genome Systems, Inc.  
(<http://www.genomesystems.com>) or Research Genetics, Inc.  
(<http://www.resgen.com>); or from Pieter de Jong.

VECTOR: PCYPAC2  
NEIGHBORING SEQUENCE INFORMATION:  
Base positions 27488 to 27520 of RP3-414A15 is represented by a  
single chemistry.

The clone sequenced to the left is RP4-693M11, 200 bp overlap; the  
clone sequenced to the right is RP1-240K6, 200 bp overlap. Actual  
start of this clone is at base position 197 of RP3-414A15; actual  
end is at base position 16578 of RP3-414A15.

FEATURES  
Location/Qualifiers

## SOURCE

1.166774  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/map="14q24.3"  
/clone="RP3-414A15"  
/clone\_1lb="RPT-3"  
139..428  
/rpt\_family="Alu"  
repeat\_region  
197..868  
/db\_xref="GI:3168733"  
579..607  
/rpt\_family="AT\_Rich"  
misc\_feature  
595..607  
/note="similar to EST AA652017 (MID:g2583669) ns47n02.sl"  
709..802  
/rpt\_family="U6"  
repeat\_region  
907..1206  
/rpt\_family="Alu"  
1793..2473  
/rpt\_family="L1"  
repeat\_region

repeat\_region 2474..2773  
/rpt\_family="Alu"  
repeat\_region 2774..2857  
/rpt\_family="L1"  
repeat\_region 2891..3148  
/rpt\_family="Alu"  
repeat\_region 3285..3403  
/rpt\_family="Alu"  
repeat\_region 3412..3455  
/rpt\_family="AT\_Rich"  
repeat\_region 3552..3595  
/rpt\_family="L2"  
repeat\_region 3643..3933  
/rpt\_family="Alu"  
repeat\_region 4298..4334  
/rpt\_family="U2"  
repeat\_region 4473..4493  
/rpt\_family="AT\_Rich"  
repeat\_region 4722..5036  
/rpt\_family="Alu"  
repeat\_region 5124..5159  
/rpt\_family="AT\_Rich"  
repeat\_region 5161..5234  
/rpt\_family="L1"  
repeat\_region 5170..5400  
/rpt\_family="L1"  
repeat\_region 5526..5559  
/rpt\_family="(A)n"  
repeat\_region 5562..5606  
/rpt\_family="(TATC)n"  
repeat\_region 5607..5891  
/rpt\_family="Alu"  
repeat\_region 6147..6224  
/rpt\_family="MER1\_type"  
repeat\_region 6423..6515  
/rpt\_family="MER1\_type"  
repeat\_region 6626..6710  
/rpt\_family="MER4\_group"  
repeat\_region 6711..7001  
/rpt\_family="Alu"  
repeat\_region 7002..7028  
/rpt\_family="(TTAAA)n"  
repeat\_region 7029..7423  
/rpt\_family="MER4\_group"  
repeat\_region 7494..7791  
/rpt\_family="L1"  
repeat\_region 7793..8092  
/rpt\_family="Alu"  
repeat\_region 8100..8332  
/rpt\_family="Alu"  
repeat\_region 8372..8411  
/rpt\_family="L1"  
repeat\_region 8412..8710  
/rpt\_family="Alu"  
repeat\_region 8711..9087  
/rpt\_family="L1"  
repeat\_region 9088..9219  
/rpt\_family="Alu"  
repeat\_region 9220..9427  
/rpt\_family="L1"  
repeat\_region 9428..9560  
/rpt\_family="Alu"  
repeat\_region 9561..9592  
/rpt\_family="L1"  
repeat\_region 9593..9726  
/rpt\_family="Alu"  
repeat\_region 9727..9806  
/rpt\_family="L1"  
repeat\_region 9809..9948  
/rpt\_family="L2"  
repeat\_region 9949..10079  
/rpt\_family="MER2\_type"  
repeat\_region 10080..10365







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2001, 19:04:36 ; Search time 1566.45 Seconds  
(without alignments)  
197.488 Million cell updates/sec

Title: US-09-675-650-4  
Perfect score: 20  
Sequence: 1 gagtaggaagatagaacg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_om:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: em\_ba1:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rod:\*  
22: em\_htg\_hum1:\*  
23: em\_htg\_hum2:\*  
24: em\_htg\_hum3:\*  
25: em\_htg\_hum4:\*  
26: em\_htg\_hum5:\*  
27: em\_htg\_hum6:\*  
28: em\_htg\_hum7:\*  
29: em\_htg\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_other:\*  
33: em\_htg\_rod:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_om:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_rod:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_vl:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_vl1:\*  
59: gb\_vl2:\*  
60: gb\_htg1:\*  
61: gb\_htg2:\*  
62: gb\_htg3:\*  
63: gb\_htg4:\*  
64: gb\_htg5:\*  
65: gb\_htg6:\*  
66: gb\_htg7:\*  
67: gb\_htg8:\*  
68: gb\_htg9:\*  
69: gb\_htg10:\*  
70: gb\_htg11:\*  
71: gb\_htg12:\*  
72: gb\_htg13:\*  
73: gb\_htg14:\*  
74: gb\_htg15:\*  
75: gb\_htg16:\*  
76: gb\_htg17:\*  
77: gb\_htg18:\*  
78: gb\_htg19:\*  
79: gb\_htg20:\*  
80: gb\_htg21:\*  
81: gb\_htg22:\*  
82: gb\_htg23:\*  
83: gb\_htg24:\*  
84: gb\_htg25:\*  
85: gb\_pr1:\*  
86: gb\_pr2:\*  
87: gb\_pr3:\*  
88: gb\_pr4:\*  
89: gb\_pr5:\*  
90: gb\_pr6:\*  
91: gb\_pr7:\*  
92: gb\_pr8:\*  
93: gb\_pr9:\*  
94: gb\_rod1:\*  
95: gb\_rod2:\*  
96: gb\_in4:\*  
97: gb\_pr10:\*  
98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	20	100.0	5435	88	AF103908 Homo sapi
c 2	20	100.0	164371	80	AL390239 Homo sapi
c 3	20	100.0	173831	80	AL359314 Homo sapi
c 4	20	100.0	267581	80	AL358573 Homo sapi
c 5	17.4	87.0	130244	63	AC013518 Homo sapi
c 6	17.4	87.0	130336	92	HS17K7 Human DNA
c 7	17.4	87.0	165525	88	AC022413 Homo sapi
c 8	17.4	87.0	176040	70	AC027548 Homo sapi

C	9	17.4	87.0	184494	77	AC087189	Homo sapi
C	10	17.4	87.0	188433	81	AL450391	Mus muscu
C	11	17.4	87.0	196840	67	AC023255	Homo sapi
C	12	17.4	87.0	197663	87	AC009245	Human Chr
C	13	17.4	87.0	210859	81	AL513347	Mus muscu
C	14	17.4	85.0	191549	79	AL158210	Homo sapi
C	15	16.8	84.0	107271	9	A96152	Sequence 18
C	16	16.8	84.0	107271	14	AT98M16	AL390921 Arabidops
C	17	16.8	84.0	125077	84	HS162E17	AL160492 Homo sapi
C	18	16.8	84.0	152036	79	AL161776	AL161776 Homo sapi
C	19	16.8	84.0	153783	90	AL157937	Human DNA
C	20	16.8	84.0	156002	71	AC036110	Homo sapi
C	21	16.8	84.0	170720	87	AC009309	Homo sapi
C	22	16.8	84.0	185573	61	AC010994	Drosophila
C	23	16.8	84.0	192104	62	AC011141	Homo sapi
C	24	16.8	84.0	192199	68	AC021810	Homo sapi
C	25	16.8	84.0	196470	66	AC024382	Homo sapi
C	26	16.8	84.0	199905	79	AL354709	Homo sapi
C	27	16.8	84.0	200573	66	AC021629	Mus muscu
C	28	16.8	84.0	209285	63	AC013517	Homo sapi
C	29	16.8	84.0	213112	81	AL513543	Homo sapi
C	30	16.8	84.0	214894	69	AC025751	Mus muscu
C	31	16.8	84.0	231464	89	AF111168	AF111168 Homo sapi
C	32	16.4	82.0	1018	72	AC054642	Giardia 1
C	33	16.4	82.0	1037	71	AC034447	Giardia 1
C	34	16.4	82.0	2911	94	AF155511	Mus muscu
C	35	16.4	82.0	56163	87	AC010126	Homo sapi
C	36	16.4	82.0	83511	12	AB013389	Arabidops
C	37	16.4	82.0	89292	91	AP001252	AP001252 Homo sapi
C	38	16.4	82.0	100061	64	AC016551	Homo sapi
C	39	16.4	82.0	107172	92	HS737M10	Human DNA
C	40	16.4	82.0	136570	64	AC016555	Homo sapi
C	41	16.4	82.0	143811	63	AC013617	Homo sapi
C	42	16.4	82.0	146596	86	AC007631	Genomic s
C	43	16.4	82.0	155313	62	AC011091	Homo sapi
C	44	16.4	82.0	155342	77	AC084269	Rattus no
C	45	16.4	82.0	155827	84	CNS05TCT	AL355099 Homo sapi

## ALIGNMENTS

RESULT 1  
AF103908 5435 bp DNA PRI 14-AUG-2000  
LOCUS Homo sapiens non-coding RNA DD3 gene, exons 2, 3, and 4.  
DEFINITION AF103908  
ACCESSION AF103908.1 GI:6165974  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 5435)  
AUTHORS Bussemakers,M.J., Van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,  
Karthaus,H.F., Schalken,J.A., Debruyne,F.M., Ru,N. and Isaacs,W.B.  
DD3: a new prostate-specific gene, highly overexpressed in prostate  
cancer  
JOURNAL Cancer Res. 59 (23), 5975-5979 (1999)  
MEDLINE 20072260  
PUBMED 10606244  
2 (bases 1 to 5435)  
AUTHORS Bussemakers,M.J.G., Van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,  
Karthaus,H.F.M., Schalken,J.A., Debruyne,F.M.J., Ru,N. and  
Isaacs,W.B.  
TITLE Direct Submission  
JOURNAL Submitted (28-OCT-1998) Urology Research Laboratory, University  
Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands  
FEATURES  
SOURCE 1..5435  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="9"

repeat\_region 1..78 /map="9q21-q22"  
exon 533..697 /rpl\_family="Alu"  
repeat\_region 1035..1294 /rpl\_type-dispersed  
mRNA join(1571..1753,1981..5435) /note="transcript III"  
mRNA join(1571..1753,1981..2517) /product="non-coding RNA DD3"  
mRNA join(1571..1753,1981..3579) /product="non-coding RNA DD3"  
mRNA join(1571..1753,1981..3579) /note="transcript (major) II"  
exon 1571..1753 /product="non-coding RNA DD3"  
exon /number=3  
exon /number=4  
exon /number=4  
polyA\_signal 2495..2499 /note="transcript I"  
polyA\_signal 2517 /note="transcript I"  
polyA\_signal 3533..3558 /note="transcript (major) II"  
polyA\_signal 3579 /note="transcript (major) II"  
repeat\_region 5268..5423 /note="LINE"  
polyA\_signal /rpl\_family="L1"  
polyA\_signal /rpl\_type-dispersed  
polyA\_signal 5435 /note="transcript III"  
BASE COUNT 1546 a 1188 c 1036 g 1654 t 11 others  
ORIGIN

Query Match 100.0%; Score 20; DB 88; Length 5435;  
Best Local Similarity 100.0%; Pred. No. 6.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 gagtagaaggatagaacg 20  
Db 1838 GAGTAGAAGGATGTAACG 1819  
|||||

RESULT 2  
AL390239/c 164371 bp DNA HMG 09-MAR-2001  
LOCUS Homo sapiens chromosome 9 clone RP11-58J3, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, 4 unordered pieces.  
ACCESSION AL390239  
VERSION AL390239.11 GI:13274794  
KEYWORDS HMG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 164371)  
AUTHORS Burton,J.  
TITLE Direct Submission  
JOURNAL Submitted (05-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Mar 12, 2001 this sequence version replaced g1:13273805.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: ba58j3  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 16337 bases at least Q40  
Consensus quality: 163481 bases at least Q30  
Consensus quality: 163577 bases at least Q20  
Insert size: 164071; sum-of-contigs  
Insert size: 166918; 4.2% error; agarose-fp  
Quality coverage: 8.78x in Q20 bases; sum-of-contigs Quality  
coverage: 9.07x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 20240: contig of 20240 bp in length  
\* 20241 20340: gap of 100 bp  
\* 20341 116100: contig of 95760 bp in length  
\* 116101 116200: gap of 100 bp  
\* 116201 118243: contig of 2043 bp in length  
\* 118244 118343: gap of 100 bp  
\* 118344 164371: contig of 46028 bp in length.  
Location/Qualifiers  
1. 164371  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/clone="RP11-58J3"  
/clone\_1lb="RP11-11.1"  
1. 20240  
/note="assembly-fragment:00838  
fragment\_chain:1  
clone\_end:SP6  
vector\_side:left"  
20341. 116100  
/note="assembly-fragment:01998  
fragment\_chain:1"  
116201. 118243  
/note="assembly-fragment:00916  
fragment\_chain:1"  
118344. 164371  
/note="assembly-fragment:00334  
fragment\_chain:1  
clone\_end:T7  
vector\_side:right"  
50254 a 34905 c 32663 g 46249 t 300 others  
BASE COUNT  
ORIGIN  
Query Match 100.0%; Score 20; DB 80; Length 164371;  
Best Local Similarity 100.0%; Pred. No. 8.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 gagtaggaagatagaacg 20  
|||||  
Db 33177 GAGTAGGAAGATAGAAACG 33158  
-----  
RESULT 3  
AL359314 173831 bp DNA HTG 08-APR-2001  
LOCUS AL359314 Homo sapiens chromosome 9 clone RP11-108L4, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*; 3 unordered pieces.  
ACCESSION AL359314  
VERSION AL359314.12 GI:13396560

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS 1 (bases 1 to 173831)  
TITLE (bases 1 to 173831)  
JOURNAL Leongamornlert,D.  
Direct Submission  
Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
On Mar 20, 2001 this sequence version replaced gi:13277120.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: ba108l4  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 173545 bases at least Q40  
Consensus quality: 173599 bases at least Q30  
Consensus quality: 173621 bases at least Q20  
Insert size: 173631; sum-of-contigs  
Insert size: 172123; 10.0% error; agarose-fp  
Quality coverage: 10.35x in Q20 bases; sum-of-contigs Quality  
coverage: 10.50x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 50595: contig of 50595 bp in length  
\* 50596 50695: gap of 100 bp  
\* 50696 157636: contig of 106941 bp in length  
\* 157637 157736: gap of 100 bp  
\* 157737 173831: contig of 16095 bp in length.  
Location/Qualifiers  
1. 173831  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/clone="RP11-108L4"  
/clone\_1lb="RP11-11.1"  
1. 50595  
/note="assembly-fragment:04137  
fragment\_chain:1"  
50696. 157636  
/note="assembly-fragment:04236  
fragment\_chain:1"  
157737. 173831  
/note="assembly-fragment:01656"  
52107 a 36469 c 35018 g 50037 t 200 others  
BASE COUNT  
ORIGIN  
Query Match 100.0%; Score 20; DB 80; Length 173831;  
Best Local Similarity 100.0%; Pred. No. 8.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 gagtaggaagatagaacg 20  
|||||  
Db 156098 GAGTAGGAAGATAGAAACG 156079  
-----  
RESULT 4



```

/note="assembly_fragment:00223
fragment_chain:1"
misc_feature
/note="assembly_fragment:00318
fragment_chain:1"
misc_feature
/note="assembly_fragment:00224
fragment_chain:2"
misc_feature
/note="assembly_fragment:00350
fragment_chain:2"
misc_feature
/note="assembly_fragment:01948
fragment_chain:3"
misc_feature
/note="assembly_fragment:02191
fragment_chain:3"
misc_feature
/note="assembly_fragment:00010"
99900..102117
misc_feature
/note="assembly_fragment:00672"
102218..109330
misc_feature
/note="assembly_fragment:01326"
109431..112187
misc_feature
/note="assembly_fragment:01498"
112288..116407
misc_feature
/note="assembly_fragment:01607"
116508..119386
misc_feature
/note="assembly_fragment:01899"
119487..121869
misc_feature
/note="assembly_fragment:02068"
121970..125546
misc_feature
/note="assembly_fragment:02469"
125647..128990
misc_feature
/note="assembly_fragment:02502"
129091..134055
misc_feature
/note="assembly_fragment:02669"
134156..138314
misc_feature
/note="assembly_fragment:02757"
138415..140612
misc_feature
/note="assembly_fragment:02776"
140713..143296
misc_feature
/note="assembly_fragment:03127"
143397..148113
misc_feature
/note="assembly_fragment:03634"
148214..151009
misc_feature
/note="assembly_fragment:03898"
151110..156616
misc_feature
/note="assembly_fragment:04220"
156717..158763
misc_feature
/note="assembly_fragment:04276"
158864..162129
misc_feature
/note="assembly_fragment:04418"
162230..164443
misc_feature
/note="assembly_fragment:04428"

Query Match      100.0%  Score 20;  DB 80;  Length 267581;
Best Local Similarity 100.0%;  Pred. No. 9;
Matches 20;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Oy      1  gagtaggaagatagaacacg 20
        |||
Db 166039 GAGTAGGAAGATAGAAACG 166058

RESULT  5
LOCUS   AC013518 130244 bp  DNA  HTG  11-APR-2000
DEFINITION Homo sapiens clone RP11-115N3, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC013518.2 GI:7533963
VERSION   AC013518.2 GI:7533963
KEYWORDS  HTG; HTGS_PHASE0.
SOURCE    human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 130244)
  Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  Homo sapiens chromosome, clone RP11-115N3
  Unpublished
  2 (bases 1 to 130244)
  Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
  Baldwin,J., Barua,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
  Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
  Cooke,P., Dearlano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
  Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
  Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
  Howland,J.C., Johnson,R., Jones,C., Kann,L., Karats,A., Klein,J.,
  Lehoczkzy,J., Lien,C., Locke,K., Macdonald,P., Margulis,N.,
  McEwan,P., McGuirk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
  Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
  Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
  Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
  Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,T., Wu,X.,
  Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  Direct Submission
  Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  On Apr 11, 2000 this sequence version replaced gi:6403763.
  All repeats were identified using RepeatMasker:
  Smit, A.F.A. & Green, P. (1996-1997)
  http://ftp.genome.washington.edu/RM/RepeatMasker.html
  ----- Genome Center
  Center: Whitehead Institute/ MIT Center for Genome Research
  Center code: WIBR
  Web site: http://www-seq.wi.mit.edu
  Contact: sequence_submissions@genome.wi.mit.edu
  ----- Project Information
  Center project name: L1903
  Center clone name: 115_N3

NOTE: This record contains 151 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
  1
  729 828: gap of 100 bp
  829 1582: contig of 754 bp in length
  1583 1682: gap of 100 bp
  1683 2482: contig of 800 bp in length
  2483 2582: gap of 100 bp
  2583 3369: contig of 787 bp in length
  3370 3469: gap of 100 bp
  3470 4225: contig of 756 bp in length
  4226 4325: gap of 100 bp
  4326 5088: contig of 763 bp in length
  5089 5188: gap of 100 bp
  5189 5950: contig of 762 bp in length
  5951 6050: gap of 100 bp
  6051 6818: contig of 768 bp in length
  6819 6918: gap of 100 bp
  6919 7686: contig of 768 bp in length
  7687 7786: gap of 100 bp
  7787 8564: contig of 778 bp in length
  8565 8664: gap of 100 bp
  8665 9409: contig of 745 bp in length
  9410 9509: gap of 100 bp
  9510 10263: contig of 754 bp in length
  10264 10363: gap of 100 bp

```

10364 11149: contig of 786 bp in length  
\* 11150 11249: gap of 100 bp  
\* 11250 12019: contig of 770 bp in length  
\* 12020 12119: gap of 100 bp  
\* 12120 12860: contig of 741 bp in length  
\* 12861 12960: gap of 100 bp  
\* 12961 13712: contig of 752 bp in length  
\* 13713 13812: gap of 100 bp  
\* 13813 14574: contig of 762 bp in length  
\* 14575 14674: gap of 100 bp  
\* 14675 15443: contig of 769 bp in length  
\* 15444 15543: gap of 100 bp  
\* 15544 16289: contig of 746 bp in length  
\* 16290 16389: gap of 100 bp  
\* 16390 17170: contig of 781 bp in length  
\* 17171 17270: gap of 100 bp  
\* 17271 18032: contig of 762 bp in length  
\* 18033 18132: gap of 100 bp  
\* 18133 18903: contig of 771 bp in length  
\* 18904 19003: gap of 100 bp  
\* 19004 19795: contig of 792 bp in length  
\* 19796 19895: gap of 100 bp  
\* 19896 20654: contig of 759 bp in length  
\* 20655 20734: gap of 100 bp  
\* 20735 21498: contig of 744 bp in length  
\* 21499 21598: gap of 100 bp  
\* 21599 22366: contig of 768 bp in length  
\* 22367 22466: gap of 100 bp  
\* 22467 23232: contig of 766 bp in length  
\* 23233 23332: gap of 100 bp  
\* 23333 24101: contig of 769 bp in length  
\* 24102 24201: gap of 100 bp  
\* 24202 24911: contig of 710 bp in length  
\* 24912 25011: gap of 100 bp  
\* 25012 25793: contig of 782 bp in length  
\* 25794 25893: gap of 100 bp  
\* 25894 26674: contig of 781 bp in length  
\* 26675 26774: gap of 100 bp  
\* 26775 27524: contig of 750 bp in length  
\* 27525 27624: gap of 100 bp  
\* 27625 28401: contig of 777 bp in length  
\* 28402 28501: gap of 100 bp  
\* 28502 29257: contig of 756 bp in length  
\* 29258 29357: gap of 100 bp  
\* 29358 30136: contig of 779 bp in length  
\* 30137 30236: gap of 100 bp  
\* 30237 30998: contig of 762 bp in length  
\* 30999 31098: gap of 100 bp  
\* 31099 31858: contig of 760 bp in length  
\* 31859 31958: gap of 100 bp  
\* 31959 32741: contig of 783 bp in length  
\* 32742 32841: gap of 100 bp  
\* 32842 33598: contig of 758 bp in length  
\* 33600 33699: gap of 100 bp  
\* 33700 34480: contig of 781 bp in length  
\* 34481 34580: gap of 100 bp  
\* 34581 35344: contig of 764 bp in length  
\* 35345 35444: gap of 100 bp  
\* 35445 36202: contig of 758 bp in length  
\* 36203 36302: gap of 100 bp  
\* 36303 37026: contig of 724 bp in length  
\* 37027 37126: gap of 100 bp  
\* 37127 37910: contig of 784 bp in length  
\* 37911 38010: gap of 100 bp  
\* 38011 38772: contig of 762 bp in length  
\* 38773 38872: gap of 100 bp  
\* 38873 39647: contig of 775 bp in length  
\* 39648 39747: gap of 100 bp  
\* 39748 40514: contig of 767 bp in length  
\* 40515 40614: gap of 100 bp  
\* 40615 41379: contig of 765 bp in length  
\* 41380 41479: gap of 100 bp  
\* 41480 42236: contig of 757 bp in length

42237 42336: gap of 100 bp  
\* 42337 43080: contig of 744 bp in length  
\* 43081 43180: gap of 100 bp  
\* 43181 43962: contig of 782 bp in length  
\* 43963 44062: gap of 100 bp  
\* 44063 44837: contig of 775 bp in length  
\* 44838 44937: gap of 100 bp  
\* 44938 45723: contig of 786 bp in length  
\* 45724 45823: gap of 100 bp  
\* 45824 46607: contig of 784 bp in length  
\* 46608 46707: gap of 100 bp  
\* 46708 47493: contig of 786 bp in length  
\* 47494 47593: gap of 100 bp  
\* 47594 48327: contig of 734 bp in length  
\* 48328 48427: gap of 100 bp  
\* 48428 49171: contig of 744 bp in length  
\* 49172 49271: gap of 100 bp  
\* 49272 50018: contig of 747 bp in length  
\* 50019 50118: gap of 100 bp  
\* 50119 50886: contig of 768 bp in length  
\* 50887 50986: gap of 100 bp  
\* 50987 51755: contig of 769 bp in length  
\* 51756 51855: gap of 100 bp  
\* 51856 52633: contig of 778 bp in length  
\* 52634 52733: gap of 100 bp  
\* 52734 53508: contig of 775 bp in length  
\* 53509 53608: gap of 100 bp  
\* 53609 54371: contig of 763 bp in length  
\* 54372 54471: gap of 100 bp  
\* 54472 55232: contig of 761 bp in length  
\* 55233 55332: gap of 100 bp  
\* 55333 56121: contig of 789 bp in length  
\* 56122 56221: gap of 100 bp  
\* 56222 56972: contig of 751 bp in length  
\* 56973 57072: gap of 100 bp  
\* 57073 57848: contig of 776 bp in length  
\* 57849 57948: gap of 100 bp  
\* 57949 58698: contig of 750 bp in length  
\* 58699 58798: gap of 100 bp  
\* 58799 59568: contig of 770 bp in length  
\* 59569 59668: gap of 100 bp  
\* 59669 60436: contig of 768 bp in length  
\* 60437 60536: gap of 100 bp  
\* 60537 61302: contig of 766 bp in length  
\* 61303 61402: gap of 100 bp  
\* 61403 62167: contig of 765 bp in length  
\* 62168 62267: gap of 100 bp  
\* 62268 63043: contig of 782 bp in length

Query Match 87.0%; Score 17.4; DB 63; Length 130244;  
Best Local Similarity 94.7%; Pred. No. 1.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gactaggaagatagaac 19  
Db 34364 GAGTAGGAAGGAAGAAGC 34346  
|||||

## RESULT 6

HS17K7 130336 bp DNA PRI 07-FEB-2000  
LOCUS Human DNA sequence from clone RPI-17K7 on chromosome 11p13.  
DEFINITION Contains three novel genes, ESTs, STSS and GSSs, complete sequence.

ACCESSION AL035078  
AL035078.32 GI:5805137  
KEYWORDS HTG.

ORGANISM human.  
SOURCE

REFERENCE 1 (bases 1 to 130336)  
AUTHORS Donnelly/S.  
TITLE Direct Submission



## JOURNAL

Submitted (05-JAN-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk

## COMMENT

On Aug 30, 1999 this sequence version replaced g1:5804869. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence was generated by the Sanger Centre from a human chromosome 11 bacterial clone contig constructed by Niederfuhr, A. et al.

<http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=979076&form=6&db=m&dopt=b>

RP1-17K7 is from the library RP1-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pCYPAC2

This sequence is the entire insert of clone RP1-17K7 The true right end of clone RP1-65P5 is at 20899 in this sequence.

## FEATURES

## Source

1..130336  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="p13"  
/clone="RP1-17K7"  
1..73  
/note="AluSp/q repeat: matches 234. .306 of consensus"  
3558..3754  
/note="MER58A repeat: matches 1. .224 of consensus"  
4589..4624  
/note="18 copies 2 mer tt 88% conserved"  
5389..5506  
/note="MER91A repeat: matches 46. .170 of consensus"  
7702..7889  
/note="MER30 repeat: matches 1. .199 of consensus"  
10794..10869  
/note="19 copies 4 mer gata 68% conserved"  
11304..11601  
/note="AluSc repeat: matches 2. .299 of consensus"  
11854..11906  
/note="MLT2CA repeat: matches 455. .508 of consensus"  
11909..11948  
/note="10 copies 4 mer atcc 87% conserved"  
11982..12017  
/note="9 copies 4 mer catc 83% conserved"  
12084..12455  
/note="MLT2CA repeat: matches 1. .444 of consensus"  
13610..13846  
/note="L1PA15 repeat: matches 5653. .5901 of consensus"  
14161..14434  
/note="L1PA15 repeat: matches 5901. .6156 of consensus"  
14517..14693  
/note="match: GSS: Em:A0830841"  
/complement(14558..14998)  
/note="match: GSS: Em:A0321591"  
14999..15026  
/note="7 copies 4 mer caca 92% conserved"  
15027..15325  
/note="L1ME3A repeat: matches 5813. .6112 of consensus"

repeat\_region 18920..18997  
/note="AluS repeat: matches 3. .80 of consensus"  
repeat\_region 20469..20856  
/note="MSRD repeat: matches 1. .394 of consensus"  
repeat\_region 20889..20928  
/note="20 copies 2 mer tg 80% conserved"  
mRNA join(<1267..21380,21558..>21684)  
/gene="dJ17K7.1"  
/note="match: ESTs: Em:AA83698"  
/evidence=not\_experimental  
/product="dJ17K7.1 (novel protein)"  
gene join(21267..21380,21558..21684)  
/gene="dJ17K7.1"  
repeat\_region 22693..22772  
/note="20 copies 4 mer ttct 68% conserved"  
repeat\_region 23117..23220  
/note="26 copies 4 mer cctt 91% conserved"  
repeat\_region 23911..24131  
/note="MER58A repeat: matches 3. .224 of consensus"  
25132..25463  
/note="MER2 repeat: matches 1. .345 of consensus"  
27272..27657  
/note="match: GSS: Em:A0560386"  
/complement(join(27441..27646,28324..28356,29647..>29713))  
/gene="dJ17K7.2"  
/evidence=not\_experimental  
/product="dJ17K7.2 (novel protein)"  
gene complement(27441..29713)  
/gene="dJ17K7.2"  
polyA\_site complement(27441)  
/gene="dJ17K7.2"  
polyA\_signal complement(27453..27458)  
/gene="dJ17K7.2"  
CDS complement(join(27622..27646,28324..28356,29647..>29713))  
/gene="dJ17K7.2"  
/note="partially supported by FGENES"  
/codon\_start=3  
/evidence=not\_experimental  
/product="dJ17K7.2 (novel protein)"  
/protein\_id="CAB72348.1"  
/db\_xref="GI:6957461"  
/translation="ILYVNLSSVEMQRFGRVHRSKIYAKILLPREILDDKC"  
repeat\_region 27806..28292  
/note="LRR40A repeat: matches 1. .501 of consensus"  
28955..29060  
/note="MER69 repeat: matches 1. .108 of consensus"  
30874..31250  
/note="match: GSS: Em:A032621"  
31251..31418  
/note="L1MA2 repeat: matches 6141. .6308 of consensus"  
31531..31727  
/note="L1PA7 repeat: matches 5949. .6145 of consensus"  
mRNA join(<33244..33419,34150..34458)  
/gene="dJ17K7.3"  
/note="isoform 2"  
match: ESTs: Em:A1027739 Em:AA725753 Em:AA921932  
Em:AA926690 Em:AA905751  
/evidence=not\_experimental  
/product="dJ17K7.3 (novel protein)"  
mRNA join(<33244..33419,37531..>37868)  
/gene="dJ17K7.3"  
/note="isoform 1"  
match: ESTs: Em:AA927867 Em:AA904797 Em:AA301105  
Em:AA926677 Em:AA883570 Em:AA904519 Em:A1026743  
Em:AA812900 Em:AA910052 Em:AA962351  
/evidence=not\_experimental  
/product="dJ17K7.3 (novel protein)"  
gene 33244..37868  
/gene="dJ17K7.3"  
repeat\_region 33961..34197  
/note="MLT1G repeat: matches 1. .251 of consensus"  
polyA\_signal 34438..34443

	polyA_site	/gene="dj17K7.3"	34458	
	repeat_region	/gene="dj17K7.3"	34943..34980	
	repeat_region	/note="19 copies 2 mer aa 81% conserved"	34950..34981	
	mRNA	/note="8 copies 4 mer aac 87% conserved"	join(c35606..c35675,37531..37868)	
		/gene="dj17K7.3"		
		/note="isoform 3		
		match: ESTs: Em:A1215480"		
		/evidence=not_experimental		
		/product="dj17K7.3 (novel protein)"	37247..37422	
	repeat_region	/note="FRAM repeat: matches 1..176 of consensus"	37795..37832	
	repeat_region	/note="19 copies 2 mer aa 100% conserved"	37846..37851	
	polyA_signal	/gene="dj17K7.3"	37868	
	polyA_site	/gene="dj17K7.3"	37884..37935	
	repeat_region	/note="13 copies 4 mer ttgt 80% conserved"	37885..37936	
	repeat_region	/note="26 copies 2 mer gt 82% conserved"	38937..39020	
	repeat_region	/note="LIMB7 repeat: matches 5534..5618 of consensus"	39399..39919	
	repeat_region	/note="LIMB7 repeat: matches 5614..6158 of consensus"	41155..41284	
	misc_feature	/note="MER21B repeat: matches 12..135 of consensus"	complement(43633..44023)	
	repeat_region	/note="match: STRS: Em:267040"	43704..43761	
	repeat_region	/note="29 copies 2 mer tg 94% conserved"	43707..43762	
	misc_feature	/note="14 copies 4 mer gtgt 94% conserved"	complement(44739..45373)	
	repeat_region	/note="match: GSS: Em:AQ319144"	47762..47783	
		/note="11 copies 2 mer ca 100% conserved"		
Query Match	87.0%; Score 17.4; DB 92;	Length 130336;		
Best Local Similarity	94.7%; Pred NO. 1.9e-02;			
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0				
QY	1 gagtggaggaatagaac 19			
Db	68214 GAGTGGGAAGGTACGAAAC 68196			
RESULT 7				
LOCUS	AC022413 166525 bp DNA PRI 25-JAN-2001			
DEFINITION	Homo sapiens chromosome 5 clone CTC-205M6, complete sequence.			
ACCESSION	AC022413			
VERSION	AC022413.4 GI:12484317			
KEYWORDS	HTG.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE				
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.			
TITLE	Direct Submission			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 166525)			
AUTHORS	DOE Joint Genome Institute.			
TITLE	Direct Submission			
JOURNAL	Submitted (03-FEB-2000) Production Sequencing Facility, DOE Joint			
	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
	3 (bases 1 to 166525)			
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.			

TITLE	Direct Submission									
JOURNAL	Submitted (25-JAN-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA									
COMMENT	On Jan 25, 2001 this sequence version replaced gi:7711721. Draft Sequence Produced by DOE Joint Genome Institute www.jgi.doe.gov Finishing Completed at Stanford Human Genome Center www.hgsc.stanford.edu Quality: Phrap Quality >=40 99.7% of Sequence: Estimated Total Number of Errors is 0.4. SYS Content: WI-15997 G21714.									
FEATURES	Location/Qualifiers									
SOURCE	1..166525 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="5" /clone="CTC-205M6"									
BASE COUNT	47370 a 36137 c 37038 g 45980 t									
ORIGIN										
Query Match	87.0%; Score 17.4; DB 88; Length 166525;									
Best Local Similarity	94.7%; Pred. No. 1.9e+02;									
Matches 18; Conservative	0; Mismatches 1; Indels 0; Gaps 0									
Oy	1 gagtagaagaagatagaac 19 									
Db 115493	GAGTAGAGAGGATGGAAC 115475									
RESULT 8										
AC027548/c	AC027548 176040 bp DNA HTG 16-MAR-2001									
LOCUS	Homo sapiens chromosome 4 clone RP11-23K2 map 4, WORKING DRAFT									
DEFINITION	SEQUENCE, 14 unordered pieces.									
ACCESSION	AC027548									
VERSION	AC027548.3 GI:13357512									
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.									
SOURCE	human.									
ORGANISM	Homo sapiens									
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.									
AUTHORS	1 (bases 1 to 176040)									
TITLE	Bliren,B., Linton,L., Nusbaum,C. and Lander,E.									
JOURNAL	Homo sapiens chromosome 4, clone RP11-23K2									
REFERENCE	Unpublished									
AUTHORS	2 (bases 1 to 176040)									
TITLE	Bliren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedalov,F., Boguski,M.S., Bonington,D., Boulikas,A., Brown,A., Burkett,G., Campione,A., Castle,A., Choe,J., Choquette,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., DeBartolotto,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kam,L., Karitas,A., Klein,J., Larocque,K., Lamazares,R., Landers,T., Leniczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N., McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPheters,R., Melnick,J., Meneus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.									
JOURNAL	Direct Submission									
COMMENT	Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On Mar 16, 2001 this sequence version replaced gi:7658429.									

All repeats were identified using RepeatMasker:  
Smil, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RW/RepeatMasker.html

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
Center project name: L8919  
Center clone name: 23\_K-2

----- Summary Statistics  
Sequencing vector: M13; M7815; 92% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 16542 bases at least Q40  
Consensus quality: 171629 bases at least Q30  
Consensus quality: 173620 bases at least Q20  
Insert size: 171000; agarose-fp  
Insert size: 174740; sum-of-contigs  
Quality coverage: 4.8 in Q20 bases; agarose-fp  
Quality coverage: 4.7 in Q20 base.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```
1 9984: contig of 9984 bp in length
* 9985 10084: gap of 100 bp
* 10085 11360: contig of 1276 bp in length
* 11361 11460: gap of 100 bp
* 11461 12753: contig of 1293 bp in length
* 12754 12853: gap of 100 bp
* 12854 14760: contig of 1907 bp in length
* 14761 14860: gap of 100 bp
* 14861 17146: contig of 2286 bp in length
* 17147 17246: gap of 100 bp
* 17247 20168: contig of 2922 bp in length
* 20169 20268: gap of 100 bp
* 20269 23632: contig of 3364 bp in length
* 23633 23732: gap of 100 bp
* 23733 30148: contig of 6416 bp in length
* 30149 30248: gap of 100 bp
* 30249 39448: contig of 9200 bp in length
* 39449 39548: gap of 100 bp
* 39549 49092: contig of 9544 bp in length
* 49093 49192: gap of 100 bp
* 49193 62340: contig of 13148 bp in length
* 62341 62440: gap of 100 bp
* 62441 77663: contig of 15223 bp in length
* 77664 77763: gap of 100 bp
* 77764 93527: contig of 15764 bp in length
* 93528 93627: gap of 100 bp
* 93628 176040: contig of 82413 bp in length.
```

FEATURES  
source  
1. 176040  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"

misc\_feature  
1. 3984  
/clone\_lib="RPCI-11 Human Male BAC"  
/note="assembly-fragment"  
clone\_end:SP6  
vector\_side:left"  
misc\_feature  
10085..11360  
/note="assembly-fragment"  
misc\_feature  
11461..12753

misc\_feature /note="assembly-fragment"  
12854..14760  
misc\_feature /note="assembly-fragment"  
14861..17146  
misc\_feature /note="assembly-fragment"  
17247..20168  
misc\_feature /note="assembly-fragment"  
20269..23632  
misc\_feature /note="assembly-fragment"  
23733..30148  
misc\_feature /note="assembly-fragment"  
30249..39448  
misc\_feature /note="assembly-fragment"  
39549..49092  
misc\_feature /note="assembly-fragment"  
49193..62340  
misc\_feature /note="assembly-fragment"  
62441..77663  
misc\_feature /note="assembly-fragment"  
77764..93527  
misc\_feature /note="assembly-fragment"  
93628..176040  
misc\_feature /note="assembly-fragment"  
/note="assembly-fragment"

BASE COUNT 49032 a 37924 c 37537 g 50226 t 1321 others  
ORIGIN

Query Match 87.0%; Score 17.4; DB 70; Length 176040;  
Best Local Similarity 94.7%; Pred. No. 1.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gagtaggaagatagaac 19  
Db 122514 GAGTGGAGGATGAAAC 122496

RESULT 9  
AC087189/c  
LOCUS AC087189  
DEFINITION Homo sapiens chromosome 16 clone RP11-43J5, WORKING DRAFT SEQUENCE,  
26 unordered pieces.  
ACCESSION AC087189  
VERSION AC087189.1 GI:11693349  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 184494)  
DOE Joint Genome Institute.  
Unpublished  
JOURNAL 2 (bases 1 to 184494)  
DOE Joint Genome Institute.  
REFERENCE DOE Joint Genome Institute.  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Submitted (13-DEC-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 0  
Center clone name: RPCI-11\_43J5  
-----  
Summary Statistics  
Consensus quality: 132143 bases at least Q40  
Consensus quality: 152435 bases at least Q30  
Consensus quality: 163165 bases at least Q20  
Estimated insert size: 183300; agarose-fp estimation  
Estimated insert size: 181994; sum-of-contigs estimation  
Quality coverage: 3.79 in Q20 bases; agarose-fp estimation

Quality coverage: 3.82 in Q20 bases; sum-of-contigs estimation.

\* NOTE: This is a 'working draft' sequence. It currently consists of 26 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
* 1 1270: contig of 1270 bp in length
* 1271 1370: gap of unknown length
* 1371 2456: contig of 1086 bp in length
* 2457 2536: gap of unknown length
* 2537 4282: contig of 1726 bp in length
* 4283 4383: gap of unknown length
* 4384 5945: contig of 1563 bp in length
* 5946 6046: gap of unknown length
* 6047 7845: contig of 1800 bp in length
* 7846 7946: gap of unknown length
* 7947 9440: contig of 1494 bp in length
* 9441 9540: gap of unknown length
* 9541 11088: contig of 1548 bp in length
* 11089 11187: gap of unknown length
* 11188 12620: contig of 1433 bp in length
* 12621 12720: gap of unknown length
* 12721 14195: contig of 1475 bp in length
* 14196 14295: gap of unknown length
* 14296 16714: contig of 2418 bp in length
* 16715 16813: gap of unknown length
* 16814 19619: contig of 2805 bp in length
* 19620 19718: gap of unknown length
* 19719 23668: contig of 3950 bp in length
* 23669 26246: gap of unknown length
* 26247 26347: gap of unknown length
* 26348 30889: contig of 4543 bp in length
* 30890 30989: gap of unknown length
* 30990 36269: contig of 5280 bp in length
* 36270 36369: gap of unknown length
* 36370 38426: contig of 2057 bp in length
* 38427 38527: gap of unknown length
* 38528 42842: contig of 4216 bp in length
* 42843 42843: gap of unknown length
* 42844 50186: contig of 7344 bp in length
* 50187 50286: gap of unknown length
* 50287 58054: contig of 7768 bp in length
* 58055 58155: gap of unknown length
* 58156 66543: contig of 8389 bp in length
* 66544 66644: gap of unknown length
* 66645 77179: contig of 10536 bp in length
* 77180 77279: gap of unknown length
* 77280 91273: contig of 13994 bp in length
* 91274 91373: gap of unknown length
* 91374 106469: contig of 15096 bp in length
* 106470 106569: gap of unknown length
* 106570 120018: contig of 13449 bp in length
* 120019 120118: gap of unknown length
* 120119 137271: contig of 17153 bp in length
* 137272 137372: gap of unknown length
* 137373 184494: contig of 47123 bp in length.
```

FEATURES

Source

```
1. .184494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-43L5"
/cclone_lib="RP11 human BAC library 11"
BASE COUNT 53403 a 38377 c 39440 g 50374 t 2900 others
ORIGIN
```

Query Match 87.0%; Score 17.4; DB 77; Length 184494;  
Best Local Similarity 94.7%; Pred. No. 2e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 gagtaggaagatagaac 19  
|| |||||  
Db 93481 GATTAGGAAAGATAGAAC 93463

RESULT 10  
AL450391/c DNA HTG 09-MAR-2001  
LOCUS  
DEFINITION Mus musculus chromosome X clone RP23-38B5, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, in unordered pieces.  
ACCESSION AL450391 GI:13121461  
VERSION AL450391.3 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
TITLE Direct Submission  
JOURNAL Submitted (03-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
Requests: clonerequests@sanger.ac.uk  
On Feb 23, 2001 this sequence version replaced gi:12581064.  
----- Genome Center  
Center: UK Medical Research Council  
Center code: UK-MRC  
Web site: http://mrcseq.har.mrc.ac.uk  
Contact: mouseq@har.mrc.ac.uk  
----- Project Information  
Center project name: DM38B5  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: Plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 182919 bases at least Q40  
Consensus quality: 183901 bases at least Q30  
Consensus quality: 184615 bases at least Q20  
Insert size: 186433; sum-of-contigs  
Insert size: 178321; 8.6% error; agarose-IP  
Quality coverage: 6.95x in Q20 bases; sum-of-contigs Quality  
coverage: 7.45x in Q20 bases; agarose-IP

\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

FEATURES

Source

```
1. .188433
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-38B5"
/cclone_lib="RP11-23"
1. .3839
/note="assembly_fragment:01310
fragment_chain:1"
3940. .7493
/note="assembly_fragment:02284
fragment_chain:1"
7594. .12260
/note="assembly_fragment:00930
fragment_chain:1"
12361. .15763
/note="assembly_fragment:00752
fragment_chain:1"
15864. .24683
/note="assembly_fragment:02693
fragment_chain:1"
24784. .36625
/note="assembly_fragment:02604
```



```
* 11829 11928: gap of 100 bp
* 11929 13104: contig of 1176 bp in length
* 13105 13204: gap of 100 bp
* 13205 14205: contig of 1001 bp in length
* 14206 14305: gap of 100 bp
* 14306 15582: contig of 1277 bp in length
* 15583 15682: gap of 100 bp
* 15683 16720: contig of 1038 bp in length
* 16721 16820: gap of 100 bp
* 16821 17978: contig of 1158 bp in length
* 17979 18078: gap of 100 bp
* 18079 19152: contig of 1074 bp in length
* 19153 19252: gap of 100 bp
* 19253 20563: contig of 1311 bp in length
* 20564 20663: gap of 100 bp
* 20664 21761: contig of 1098 bp in length
* 21762 21861: gap of 100 bp
* 21862 22902: contig of 1041 bp in length
* 22903 23002: gap of 100 bp
* 23003 24045: contig of 1043 bp in length
* 24046 24145: gap of 100 bp
* 24146 25692: contig of 1547 bp in length
* 25693 25792: gap of 100 bp
* 25793 27574: contig of 1782 bp in length
* 27575 27674: gap of 100 bp
* 27675 29750: contig of 2076 bp in length
* 29751 29850: gap of 100 bp
* 29851 32791: contig of 2941 bp in length
* 32792 32891: gap of 100 bp
* 32892 35428: contig of 2537 bp in length
* 35429 35528: gap of 100 bp
* 35529 41425: contig of 5897 bp in length
* 41426 41525: gap of 100 bp
* 41526 47276: contig of 5751 bp in length
* 47277 47376: gap of 100 bp
* 47377 52698: contig of 5322 bp in length
* 52699 52798: gap of 100 bp
* 52799 62767: contig of 9969 bp in length
* 62768 62867: gap of 100 bp
* 62868 78028: contig of 15161 bp in length
* 78029 78128: gap of 100 bp
* 78129 94604: contig of 16476 bp in length
* 94605 94704: gap of 100 bp
* 94705 112411: contig of 17707 bp in length
* 112412 112511: gap of 100 bp
* 112512 148635: contig of 36124 bp in length
* 148636 148735: gap of 100 bp
* 148736 196840: contig of 48105 bp in length.
FEATURES
source
1. .196840
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-754A4"
/clone_id="RP11-11 Human Male BAC"
1. .1001
/note="assembly_fragment"
misc_feature
1102. .2141
/note="assembly_fragment"
misc_feature
2242. .3546
/note="assembly_fragment"
misc_feature
3647. .4739
/note="assembly_fragment"
misc_feature
4840. .5885
/note="assembly_fragment"
misc_feature
5986. .7222
/note="assembly_fragment"
misc_feature
7323. .8677
/note="assembly_fragment"
misc_feature
8778. .10491
/note="assembly_fragment"
misc_feature
10592. .11828
/note="assembly_fragment"
misc_feature
11929. .13104
```

```
/note="assembly_fragment"
misc_feature
13205. .14205
/note="assembly_fragment"
misc_feature
14306. .15582
/note="assembly_fragment"
misc_feature
15683. .16720
/note="assembly_fragment"
misc_feature
16821. .17978
/note="assembly_fragment"
misc_feature
18079. .19152
/note="assembly_fragment"
misc_feature
19253. .20563
/note="assembly_fragment"
misc_feature
20664. .21761
/note="assembly_fragment"
misc_feature
21862. .22902
/note="assembly_fragment"
misc_feature
23003. .24045
/note="assembly_fragment"
misc_feature
24146. .25692
/note="assembly_fragment"
misc_feature
25793. .27574
/note="assembly_fragment"
misc_feature
27675. .29750
/note="assembly_fragment"
misc_feature
29851. .32791
/note="assembly_fragment"
misc_feature
32892. .35428
/note="assembly_fragment"
misc_feature
clone_end:77
vector_side:right"
35529. .41425
/note="assembly_fragment"
misc_feature
41526. .47276
/note="assembly_fragment"
misc_feature
47377. .52698
/note="assembly_fragment"
misc_feature
52799. .62767
/note="assembly_fragment"
misc_feature
62868. .78028
/note="assembly_fragment"
misc_feature
78129. .94604
/note="assembly_fragment"
misc_feature
```

```
Query Match 87.0%; Score 17.4; DB 67; Length 196840;
Best Local Similarity 94.7%; Pred. No. 2e+02; 1; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gagtaggaagatagaac 19
|||||
Db 43583 GAGTAGAAGATAGAAC 43601
```

```
RESULT 12
AC009245 197663 bp DNA PRI 29-JUN-2000
LOCUS AC009245 Human Chromosome 7 clone RP11-351B12, complete sequence.
DEFINITION AC009245
ACCESSION AC009245
VERSION AC009245.10 GI:8810492
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 197663)
AUTHORS Kaul,R.K., Zhou,Y., James,R.A., Raymond,C., Haugen,E.D. and
Olson,M.V.
TITLE Large-scale Mapping and Sequencing of Human Chromosome 7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 197663)
AUTHORS Bubb,K.L., Desmarrats,C.L., Ramsey,S.A. and Hubley,R.M.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-1999) Human Genome Center, University of
```

REFERENCE Washington, Box 352145, Seattle, WA 98195, USA  
 3 (bases 1 to 197663)  
 AUTHORS Kaul, R.K., Zhou, Y., James, R.A., Raymond, C., Haugen, E.D. and Olson, M.V.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-JUN-2000) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA  
 COMMENT On Jun 29, 2000 this sequence version replaced gi:8039711.  
 ----- Genome Center -----  
 Center: University of Washington Genome Center  
 Center Code: UMGc  
 Web site: <http://www.genome.washington.edu/UMGC/>  
 Contact: [umgchgs@u.washington.edu](mailto:umgchgs@u.washington.edu)  
 ----- Project Information -----  
 Center project name: HaCh7  
 Center clone name: RP11-351B12 (d1s263)  
 ----- Summary Statistics -----  
 Sequencing vector: Plasmid; X52328; 100% of reads  
 Chemistry: Dye-primer-amerham; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 1138759 bases at least Q40  
 Consensus quality: 1528231 bases at least Q20  
 Consensus quality: 1749145 bases at least Q20  
 Insert size: 210450; 4.1% error; agarose-fp  
 Insert size: 197663; sum-of-contigs  
 Quality coverage: 8.31x in Q20 bases; agarose-fp  
 Quality coverage: 8.85x in Q20 bases; sum-of-contigs

Overlapping Sequences:  
 5': RP11-22814 (UMGC:d1s352)  
 3': RP11-2983 (UMGC:d1s300)

Sequence Quality Assessment:  
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:  
 all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.  
 Double stranded (DS) coverage: 92.9%  
 DS or two chemistry coverage: 6.82%  
 Single stranded regions: 0.238% in 5 gaps

Sequence Validation:  
 This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector. In order to accurately represent the entire circular BAC, small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

Bg11 EcoRI  
 -----  
 SeqDerMap FngPrnt SeqDerMap FngPrnt SeqDerMap FngPrnt  
 SeqDerMap FngPrnt  
 -----

5535	5377	8696	8824	14525	14394
8696	8822				
2067	2046	510	<800	3462	3515
510	<800				
4620	4587	6221	6154	1637	1583
6221	6168				
834	809	12486	12560	286	<800
12486	12543				
10311	10377	97	<800	2229	2327
97	<800				
13707	13646	4074	4231	1264	1215
4074	4202				
1356	1319	12812	12560	4514	4528
12812	12543				
7643	7323	21	<800	705	<800
21	<800				
10251	10377	1282	1331	1944	1909
1282	1336				
1951	2046	1832	1898	1202	1215
1832	1901				
628	<800	4277	4231	119	<800
4277	4202				
3987	3901	912	939	396	<800
912	940				
597	<800	3686	3650	2403	2327
3686	3639				
3588	3552	729	<800	1149	1215
729	<800				
822	809	1339	1331	1576	1583
1339	1336				
5087	5104	1040	1046	2768	2975
1040	1050				
941	919	3112	3135	1442	1417
3112	3138				
5377	5377	964	939	2969	2975
964	940				

```
-----
208      <800      1375      1331      732      <800
1375      1336      -----
156      <800      448      <800      1189      1215
448      <800      -----
784      809      4901      4916      1057      1092
4801      4897      -----
508      <800      466      <800      5203      5174
466      <800      -----
1447      1376      8786      8824      1551      1583
8786      8822      -----
1344      1319      6576      6593      16871      16417
6576      6600      -----
1787      1928      4111      4231      3995      4127
4111      4202      -----
4772      4805      5414      5404      1647      1777
5414      5398      -----
666      <800      2312      2322      1232      1215
2312      2317      -----
Query Match      87.0%; Score 17.4; DB 87; Length 197663;
Best Local Similarity 94.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy      1 gagtagagaagatagaac 19
      |||
Db 79197 GAGTAGAGAAGTAAGAAC 79215
```

```
RESULT 13
AL513347      210859 bp      DNA      HTG      07-APR-2001
LOCUS      Mus musculus chromosome X clone RP23-476D16, *** SEQUENCING IN
DEFINITION      AL513347
ACCESSION      AL513347
VERSION      HTG: HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS      house mouse.
SOURCE      Mus musculus.
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
      1 (bases 1 to 210859)
REFERENCE      Direct Submission
      Sanger Centre, Hinxton, Cambridgeshire,
AUTHORS      Submitted (06-APR-2001)
      Sanger Centre, Hinxton, Cambridgeshire,
      CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
      requests: clonerequests@sanger.ac.uk
      On Apr 8, 2001 this sequence version replaced g1:12718187.
COMMENT      -----
      Genome Center
      Center: Sanger Centre
      Center code: SC
      Web site: http://www.sanger.ac.uk
      Contact: humquery@sanger.ac.uk
      Project Information
      Center project name: Bm476D16
      -----
      Summary Statistics
```

Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 208726 bases at least Q40  
Consensus quality: 209071 bases at least Q40  
Consensus quality: 209334 bases at least Q20  
Insert size: 209659; sum-of-contigs  
Insert size: 201668; 4.4% error; agarose-fp  
Quality coverage: 13.11x in Q20 bases; sum-of-contigs Quality  
coverage: 13.98x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

## FEATURES

```
source      1. .210859
      Location/Qualifiers
misc_feature      1. .4580
      /note="assembly_fragment:05672
      fragment_chain:1"
misc_feature      4681. .8555
      /note="assembly_fragment:07375
      fragment_chain:1"
misc_feature      8656. .31018
      /note="assembly_fragment:08111
      fragment_chain:1"
misc_feature      31119. .34560
      /note="assembly_fragment:02143
      fragment_chain:1"
misc_feature      34661. .39382
      /note="assembly_fragment:06785
      fragment_chain:1"
misc_feature      39483. .47527
      /note="assembly_fragment:06336
      fragment_chain:1"
misc_feature      47628. .72453
      /note="assembly_fragment:07547
      fragment_chain:2"
misc_feature      72554. .95012
      /note="assembly_fragment:07511
      fragment_chain:2"
misc_feature      95113. .149922
      /note="assembly_fragment:02768
      fragment_chain:2"
misc_feature      150823. .160893
      /note="assembly_fragment:00710
      fragment_chain:2"
misc_feature      160994. .166350
      /note="assembly_fragment:05231
      fragment_chain:2"
misc_feature      166451. .207988
      /note="assembly_fragment:00818"
      fragment_chain:2"
misc_feature      208089. .210859
      /note="assembly_fragment:03090"
      fragment_chain:2"
BASE COUNT      63414 a 43413 c 43076 g 59755 t 1201 others
ORIGIN
```

```
Query Match      87.0%; Score 17.4; DB 81; Length 210859;
Best Local Similarity 94.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy      1 gagtagagaagatagaac 19
      |||
Db 78060 GAGTAGAGAAGTAAGAAC 78078
```

RESULT 14



AL158210/c  
 LOCUS AL158210 191549 bp DNA HTG 08-APR-2001  
 DEFINITION Homo sapiens chromosome 10 clone RP11-55418, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*; 4 unordered pieces.  
 ACCESSION AL158210  
 VERSION AL158210.10 GI:13567917  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 191549)  
 REFERENCE BURTON, J.  
 AUTHORS Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 TITLE CB10 15A, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 JOURNAL requests: clonerequests@sanger.ac.uk  
 COMMENT On Apr 9, 2001 this sequence version replaced g1:11228972.  
 ----- Genome Center  
 Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk  
 ----- Project Information  
 Center project name: ba55418  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Consensus quality: 189400 bases at least Q40  
 Consensus quality: 189884 bases at least Q30  
 Consensus quality: 190273 bases at least Q20  
 Insert size: 191249; sum-of-contrigs  
 Insert size: 170480; 7.1% error; agarose-fp  
 Quality coverage: 7.73x in Q20 bases; sum-of-contrigs quality  
 coverage: 8.72x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 63110: contig of 63110 bp in length  
 \* 63111 63210: gap of 100 bp  
 \* 63211 84560: contig of 21350 bp in length  
 \* 84561 84660: gap of 100 bp  
 \* 84661 188906: contig of 104246 bp in length  
 \* 188907 189006: gap of 100 bp  
 \* 189007 191549: contig of 2343 bp in length.  
 Location/Qualifiers  
 1. 191549  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-55418"  
 /clone\_1b="RPC1-11.2"  
 1. 63110  
 /note="assembly\_fragment:02425  
 clone\_end:SP6  
 vector\_side:left"  
 63211. 84560  
 /note="assembly\_fragment:01526  
 fragment\_chain:1"  
 84661. 188906  
 /note="assembly\_fragment:00575  
 fragment\_chain:1"  
 189007. 191549  
 /note="assembly\_fragment:00076"  
 misc\_feature  
 misc\_feature  
 misc\_feature  
 misc\_feature  
 BASE COUNT 56923 a 40957 c 38670 g 54690 t 309 others  
 ORIGIN

Query Match 85.0%; Score 17; DB 79; Length 191549;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 3 gtaggaagcataaac 19  
 |||  
 Db 88624 GTAGGAAGCATGACAC 88608  
 RESULT 15  
 A96152/c  
 LOCUS A96152 792 bp DNA PAT 07-SEP-2000  
 DEFINITION Sequence 185 from Patent WO9924578.  
 ACCESSION A96152  
 VERSION A96152.1 GI:6779968  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 792)  
 AUTHORS Pizza, M., Scarlato, V., Rappuoli, R., Grandi, G. and Masignani, V.  
 TITLE Neisserial antigens  
 JOURNAL Patent: WO 9924578-A 185 20-MAY-1999;  
 PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);  
 CHITON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)  
 FEATURES  
 source  
 1. 792  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 210 a 184 c 189 g 209 t  
 ORIGIN  
 Query Match 84.0%; Score 16.8; DB 9; Length 792;  
 Best Local Similarity 90.0%; Pred. No. 2.4e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 1 gagtaggaagcataaac 20  
 |||  
 Db 148 GAGTAGGAAGTGAAGAAACG 129

Search completed: October 16, 2001, 19:05:32  
 Job time: 4465 sec

---